My genome. So what?

Research is needed into the way individuals use their genomic information, and into protection from its abuse by others.

Human genome research has proved itself predictably unpredictable. As was widely anticipated, the speed of sequencing has escalated, the pace of linking genes to disease has quickened, and practically anyone can have their genome investigated and fed back to them in electronic format to do with it what they will. In this issue, two groups reveal individual genome sequences of a Yoruba man from Ibadan, Nigeria (see page 53), and of a Han Chinese individual (see page 60) for a cost of less than US$500,000 each — a fraction of that of the human genome's first drafts or subsequently published editions.

The age of personal genomes is here. What many promoters of genomics did not predict are the challenges that individuals face in using this information. One is the limited extent to which the genetic constitution revealed says anything about future health. The predictive value of genetic associations has fallen short of some expectations, often in dramatic ways (see page 18), and fails to augment in any meaningful way more traditional predictors for health, such as lifestyle and family history.

Another largely unpredicted outcome were the private companies that sprang up to capitalize on these genetic clues, selling individuals genotype information and predictions about health based on the incomplete information available. Yet more services are now springing up to help people make sense of the data (see page 11). It's obvious that this genetic fortune-telling will be murky and inconclusive for many years to come. What is not clear is how people will act on it. However questionably, these companies are blazing a trail that could provide insight into ways in which people interact with these sensitive personal data. Researchers should look for more ways to investigate these consumer interactions from the perspective of public health, social sciences and potential biomedical applications.

One predicted outcome of human genome sequencing was the stocking of the drained pipelines of pharmaceutical companies with drug targets in the hope of ultimately developing cures for common afflictions. Although there have been some rousing signs of success, the refrain is that the complexity of the problem requires more data and more research (see page 26). Personal genomes may have a useful role: more human sequences complete with thorough medical histories and information about the environments in which the individuals grew up and lived will be the richest source of data to understand the genetic underpinnings of disease.

But making such information easily available to researchers has predictably challenged valued principles of privacy. Current protection for research subjects is inadequate in this respect (see page 32). In the United States, the Genetic Information Nondiscrimination Act of 2008 provides safeguards against discrimination by employers and health-insurance companies but does not protect against other potential misuse, including intrusion by law-enforcement agencies. Anonymizing data is not the answer, as re-identification of anonymous data can be easy. And researchers' discretion cannot be relied on.

Notions of privacy are changing — many people seem quite willing to share information about their genomes and medical histories. But researchers could make better use of available ways of protecting the privacy of their research subjects. Certificates of confidentiality in the United States give researchers the right to refuse disclosure to any civil authority of any information that could identify a subject, and are one example of a possibly underused protection. Researchers need to collaborate with social scientists, legal experts and regulators to improve on such models, both for the current challenges to privacy that personal genomes pose and for challenges that have yet to present themselves. We can't predict everything that will happen next, but we can be prepared.

For podcast and more online extras see www.nature.com/nature/focus/personal genomes/index.html

1

Animals aren't drugs

The US Food and Drug Administration is misguided in its approach to genetically modified animals.

It is more than 25 years since Ralph Brinster and Richard Palmiter first developed genetically engineered (GE) mice, proving that recombinant DNA techniques could be used to engineer animals. It has taken the US Food and Drug Administration (FDA) nearly as long to develop guidelines laying out its regulatory approach to such animals, be they intended as pets, as living drug factories or to supply American dinner tables.

In September, the FDA finally delivered its draft guidelines, effectively laying out a detailed playbook for companies seeking the agency's seal of approval to bring to market everything from fast-growing salmon to pigs with livers engineered for human transplant. The period for public comment on this important FDA document ends on 18 November (see http://www.fda.gov/cvm/GEAnimals.htm).

It is high time that the FDA stepped in to regulate this field, in which companies such as Aqua Bounty Technologies, a small Massachusetts enterprise that has engineered a salmon that grows to marketable adult weight in 18 months instead of 30, have been undermined by the agency's slowness to act. Agency involvement will, furthermore, bring needed regulatory oversight to an enterprise that, although often promising, could in individual instances go awry with unhappy and unpredictable consequences for the animals, public health and the environment.

But the agency's regulatory approach to the issue is troubling. It has used an eyebrow-raising reading of the 1938 Federal Food, Drug and Cosmetic Act to assert its regulatory authority over GE animals. The FDA says, in effect, that these animals meet the definition of a 'drug' under the law because they contain DNA that is "intended to affect the structure or function of the body." Following from this, the guidance says that every new GE animal — with the notable exception of lab animals used in research — will be regulated as if it contains a new drug.

When a conventional drug is being assessed by the FDA, the existence and details of the application are protected under law from public scrutiny. Such protections are necessary in the highly competitive world of human pharmaceuticals. Applied to GE animals, they are much less appropriate. In essence, the agency is saying to the public 'trust us' — in the absence of evidence, for example, that it is adequately equipped to assess the potential environmental impacts of such animals. It is not just environmentalists who are raising the red flag; the National Research Council, in a 2002 report on animal biotechnology, listed "novel environmental issues" and the technological capacities of agencies like the FDA as among its "major concerns".

It is understandable that the agency is trying to pour new wine into the 70-year-old wineskin of the federal drug law; the law is the only tool at its disposal for regulating GE animals. But as Henry Miller of the Hoover Institution at Stanford University, California, noted in a recent correspondence in *Nature Biotechnology*, "When the only tool you have is a hammer, more and more problems begin to look like nails." (See http://www.nature.com/nbt/journal/v26/n2/full/nbt0208-159.html.)

More light on the process than the FDA's proposal allows is needed to build public trust and to ensure that all necessary steps are taken to avoid adverse events. The current law cannot do this. Congress should step in and produce one that does. ∎

Scientists and rights

Researchers should support new initiatives aimed at engaging them with human-rights groups.

Six foreign medics escaped the Libyan death penalty last year thanks to intense diplomacy, supported by the advocacy and decisive expertise of scientists. But the researchers' involvement was largely a matter of luck and serendipity. Science and scientists have much untapped potential to contribute to human-rights issues, but until now there have been limited efforts to systematically consolidate the interactions between science and human-rights groups. Two new initiatives of the Science and Human Rights Program of the American Association for the Advancement of Science are intended to help fill that gap.

Its "On-call" Scientists program launched last month aims to create a database of scientists who will volunteer time — be it a few days or a few months — and expertise, and human-rights organizations — including non-governmental organizations and international agencies such as the United Nations — seeking practical help or advice. (See http://oncallscientists.aaas.org/default.aspx.)

'Human rights' covers a gamut of issues, from exposing abuses to disaster relief. The range of scientific advice sought is correspondingly broad — statistical or methodological help to get a more accurate picture of conflict or ethnic cleansing, advice on water issues from hydrologists, or forensic help to document mass executions or overturn false convictions.

The service faces a steep learning curve in deciphering the diverse needs of human-rights groups, and how scientists might be able to help in ways perhaps not yet imagined. But better communication between scientists and the alphabet soup of human-rights groups — and between those groups themselves on technical issues — is long overdue.

Another welcome initiative is due in January 2009. Many learned societies, as well as academic groups such as Scholars at Risk, have a long history in upholding human rights and academic freedom — for example, defending scientists under threat from oppressive governments, using satellite imagery to expose human-rights abuses and speaking out on abuse wherever it occurs. To put such efforts on a firmer footing, American organizations are to launch the US Science and Human Rights Coalition, a forum in which scientific bodies and human-rights groups can share experiences and best practice. Given the US presidential election, the timing could not be better. For the past eight years, American human-rights groups have seen their international influence undermined by the US administration's diminishing moral authority and standing in the world. Scientists can, and should, help reinstate the fundamental principles enshrined in the Universal Declaration of Human Rights in 1948. ∎

RESEARCH HIGHLIGHTS

Famine's shadow

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0806560105 (2008)
If a starving woman becomes pregnant, her child's DNA can still bear traces of her hunger more than six decades later.

Lambert Lumey of Columbia University in New York, Bastiaan Heijmans of Leiden University Medical Center in the Netherlands and their colleagues studied the methyl groups attached to a gene called *IFG2*. They measured methylation at five points along *IFG2* in people prenatally exposed to the 1944–45 Dutch famine — when a Nazi embargo led to food rationing in the west of Holland of fewer than 700 calories a day.

Compared with same-sex siblings conceived when the same mothers had more flesh on their bones, those affected early in fetal development have less methylation on *IFG2* today, implying that their cells express it more readily.

PHYSICS
Big little things

Phys. Rev. Lett. **101,** 171805 (2008)
The top quark is roughly 40 times as massive as the second heaviest quark, the bottom. But why?

Hsin-Chia Cheng and his colleagues at the University of California, Davis, propose that top quarks may have a 'superpartner' with a spin of 1 rather than spin 0 as is usually predicted by supersymmetry theory. Spin-1 particles tend to mediate forces; the photon, for example, is a spin-1 particle and is responsible for electromagnetism. The new particle, the researchers propose, would mediate a force determining the interaction of the top quark with the Higgs boson, which putatively gives things mass.

If such a particle were to exist, it should be detectable with the Large Hadron Collider, when that is back in action.

PARASITOLOGY
The bacterial racketeer

Science **322,** 702 (2008)
Wolbachia are well known bacteria because they often kill developing males of all manner of creatures, from nematodes to crustaceans. Karyn Johnson and her colleagues from the University of Queensland in Brisbane, Australia, now report that *Wolbachia* offer fruitflies some protection against diverse and deadly RNA viruses.

They compared the survival of two strains of fruitfly infected with *Drosophila* C virus with that of the same species infected with both this virus and *Wolbachia pipientis*. The bacterium seemed to delay virus-induced mortality by the same amount of time that it delayed the accumulation of virus particles in the flies, implying a causal link. Johnson's team then tested two other viruses in the same way, and also found that *Wolbachia* delayed mortality.

ZOOLOGY
Green growth

J. Exp. Zool. doi:10.1002/jez.497 (2008)
The size of a flatfish, and of its appetite, is influenced by the colour of its environment. Akiyoshi Takahashi of Kitasato University in Iwate, Japan, and his colleagues have discovered that the barfin flounder (*Verasper moseri*), a promising species for aquaculture, grows longer and heavier if kept under green light than under blue or unfiltered light. Red light seems to stunt its growth.

The team kept adult fish of the same approximate starting size for 14 weeks, giving them as many pellets as they were willing to eat twice a day. The different wavelengths of light the fish experienced may have modified the release of melanin-concentrating hormone, an appetite stimulant, in the brain, prompting the fish to eat more, the authors say. Like many fish, this species keeps growing throughout its life.

NANOTECHNOLOGY
Future pixels

Adv. Mater. doi:10.1002/adma.200801167 (2008)
Tiny marbles, black on one side and coloured on the other, can be made by 'curing' suspensions of silica particles with an ultraviolet lamp, according to Seung-Man Yang and his colleagues at the Korea Advanced Institute of Science and Technology in Daejeon. When an electric field is applied, the marbles line up so that the black sides all face upwards, which suggests they may prove useful pigments for flexible electronic displays.

The researchers suspended a flow of carbon-black particles mixed with silica and a transparent or coloured silica flow in a resin that polymerizes under ultraviolet light. They then passed the mixture through a tiny see-through tube. The light solidified the silica and resin as balls with differently coloured regions (pictured left), each about 200 micrometres in diameter.

MOLECULAR BIOLOGY
Ubiquitous no more

Cell **135,** 462–474 (2008)
One of two processes thought to be catalysed by RNA and common to all life forms does not actually need its RNA.

The RNase P catalyst, normally made of RNA and protein, chops superfluous subunits off immature versions of tRNA

molecules, which are essential for protein synthesis. A quarter of a century ago, this catalyst's RNA component was shown to be crucial to its function in bacterial cells; since then, researchers have shown that this RNA can do the job without any help from proteins in the two other evolutionary branches of life, archaea and eukaryotes.

But Walter Rossmanith of the Medical University of Vienna and his colleagues have identified and purified the components of human mitochondrial RNase P, finding only proteins, and reconstituted its catalytic activity using just three of these.

ASTRONOMY
Hidden gems

Astrophys. J. doi:10.1086/592037 (2008)
If the remnants of the first stars were to be found, they should be in small galaxy groups — relatively common structures. Calculations by Michele Trenti, now of the University of Colorado at Boulder, and his colleagues also suggest that the earliest, brightest quasars evolved to become part of galactic groups of medium brightness.

This is at odds with current theory, which puts remnants of the first stars — born when the Universe was just 65 million years old — in the largest observable clusters in the present-day Universe. Similarly, current theory places the remnants of the brightest quasars from about 1 billion years after the Big Bang in the largest clusters.

NEUROSCIENCE
Making memories

Cell **135,** 535–548 (2008)
A protein called MyoVb may aid learning and memory by helping to strengthen connections between neurons.

Memories are thought to form through a process of 'long-term potentiation', which improves communication between neurons that fire simultaneously. This requires the transport of molecules to small spines sticking out of neurons. The spines receive electrical signals from other neurons.

Michael Ehlers of Duke University Medical Center in Durham, North Carolina, and his colleagues have discovered that MyoVb moves the vesicles that transport molecules down spines during long-term potentiation. Eliminating MyoVb levels blocked spine growth. It also stopped a type of receptor that is important for rapid communication between neurons being inserted into the spines' membranes. Chemically blocking MyoVb halted long-term potentiation in mouse brain slices.

GEOSCIENCES
Join the club

Nature Geosci. doi:10.1038/ngeo338 (2008)
Antarctica can finally be included in the list of places warmed by human activity. Nathan Gillett of the University of East Anglia, UK, and his colleagues have shown a dear human influence on temperatures at both the North and South Poles with data going back to 1900 and 1950, respectively.

They compared the available data from both poles to simulations from four climate models. The records from both poles could not be explained by natural variation or natural driving forces alone.

So far, the Intergovernmental Panel on Climate Change has said there are insufficient data to point the finger at an anthropogenic impact in Antarctica. Gillett thinks that conclusion is due for an update.

E. ROBISON/ALAMY

THEORETICAL PHYSICS
Toppling tubes

Phys. Rev. Lett. **101,** 175501 (2008)
Nanotubes made of a honeycomb arrangement of carbon atoms are famed for their strength, but Tienchong Chang of Shanghai University in China has found a chink in their armour. His calculations show that pinching a single-walled carbon nanotube at its end will cause it to collapse along its entire length. The effect is rather like toppling dominos, but in this case the electric charge along the tube, rather than gravity, drives the self-propagating collapse.

This weakness may prove to be a strength. Chang proposes new applications for nanotubes that collapse in this way, including a 'nanogun' for injecting or expelling molecules from devices.

Correction
The Research Highlight 'Twitchy details' (*Nature* **455,** 1152–1153; 2008) stated that Lin Mei is at the Johns Hopkins University School of Medicine in Baltimore, Maryland. He is in fact at the Medical College of Georgia in Augusta.

JOURNAL CLUB

Shanan Peters
University of Wisconsin–Madison

A geologist questions a grand theory.

Atmospheric oxygen concentrations are falling. Breathing is difficult. Those that can't cope are collapsing and dying with symptoms akin to altitude sickness.

This may read like the first page of a Hollywood script, but, according to the oxygen-stress hypothesis, a similar scene occurred 251 million years ago at the end-Permian mass extinction, when up to 95% of all animal species died out. Like all good prevailing hypotheses, this one makes predictions that can be tested, if only the right rocks can be found.

Enter Tyler Beatty of the University of Calgary in Alberta, Canada, and his colleagues. They recently set up camp in the remote reaches of northwestern Canada, where rocks spanning the end-Permian extinction show a shift from Permian sandy carbonates to Triassic sand and mud. They found that fossils of entire creatures are not common at the boundary, preventing taxonomic analyses, but that fossils documenting sediment disturbance by animals are (T. W. Beatty *et al. Geology* **36,** 771–774; 2008). This is fortuitous because such disturbance in marine sediments is linked to oxygen concentration. So these rocks may preserve a 'smoking gun' for an oxygen-stressed world.

However, the shallow marine sediments of the Early Triassic were pervasively burrowed by diverse organisms of the period, including large, oxygen-demanding arthropods. Only deeper-water sediments, deposited below wave-mixed surface waters, had the expected oxygen-stressed fossil traces.

This complicates the oxygen-stress story for the end-Permian mass extinction. Beatty *et al.* stop short of asking whether the end-Permian mass extinction was really caused by a massive reduction in atmospheric oxygen. But in light of their results, I am not holding my breath.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

Industry shifts focus to immunology and cancer

Cardiology and anaemia lose out in the hunt for the next pharmaceutical blockbusters.

Pfizer's cholesterol-lowering Lipitor was the world's best-selling drug in 2007.

When Wyeth Pharmaceuticals announced last week that it would cut some of its research and development (R&D) programmes in women's health, the decision seemed counter-intuitive. The pharmaceutical giant, based in New Jersey, is known for its strong work in contraceptives and hormone-replacement therapy. The decision also sounded familiar: on 30 September, the New York-based company Pfizer said it would pare down its R&D efforts as well, eliminating research programmes in nine disease areas including cholesterol — an area in which Pfizer has been a leader.

Both companies, like much of the pharmaceutical industry worldwide, are tightening their belts as they face looming competition from generic drugs, increasingly conservative drug regulators and diminishing product pipelines. Pfizer faces patent expiration in 2010 on its multibillion-dollar cholesterol-lowering drug, Lipitor

"New drugs must perform much better than those already on the market before regulators are willing to take a chance."

(atorvastatin), and Wyeth has launched a cost-cutting effort called Project Impact that aims to trim thousands of employees from its payroll.

In response to such challenges, the drug industry is shifting research away from its bread-and-butter 'primary care' products — those likely to be prescribed by a primary-care physician, once prized for their massive markets — to 'speciality' drugs prescribed by specialists in fields such as oncology and neurology. These drugs captured 45% of pharmaceutical sales in 2006, up from 39% in 2001 (M. Gudiksen, E. Fleming, L. Furstenthal and P. Ma, *Nature Rev. Drug Discov.* **7**, 563–567; 2008). "Big pharma has prided itself on being very diverse in its approach to R&D," says Kenneth Kaitin, director of the Tufts Center for the Study of Drug Development in Boston, Massachusetts. "But that's just not feasible these days in the current market conditions."

Wyeth plans to shrink the number of diseases it tackles from 55 to 27. It will instead increase its focus on cancer as well as inflammatory, metabolic and neurological disorders — fields that remain on Pfizer's R&D slate as well. "These areas represent our highest probability for success," says Pfizer spokeswoman Kristin Neese, who cites unmet medical need and potential for market growth as two selection criteria. Meanwhile, Pfizer's eliminated research programmes read like a grocery list of primary-care conditions: cholesterol, osteoporosis, gastrointestinal conditions, osteoarthritis and anaemia.

Bush may introduce environmental regulations

In its waning days, the administration of President George W. Bush may roll out a number of new environmental regulations, the effects of which could persist long after Bush leaves office on 20 January 2009.

Last week, for instance, the US Environmental Protection Agency (EPA) instituted new environmental regulations for factory farms. The EPA says that the regulations would curb the amount of nitrogen, phosphorus and sediment entering waterways, and farm operators have greeted it with cautious optimism. But environmentalists say a loophole in the rule would scale back environmental protection by effectively allowing operators to police themselves — on these and other requirements — under the Clean Water Act.

Environmentalists and public-policy watchdogs are expecting similar industry-friendly regulatory changes in the coming months. Such 'midnight regulations' have become common practice in recent decades as presidents, both Republican and Democrat, seek to leave their mark on public policy.

More than a dozen rules that could be changed are being monitored by OMB Watch, a Washington DC-based advocacy group, and others. One of these rules would make it easier for mountain-top mine operators to dump debris in streams. Others would ease air-quality restrictions on power plants operating near national parks and wilderness areas, and would make it easier for utilities to update old power plants

without triggering a requirement to install modern pollution controls.

"There are a lot of rules out there that could potentially move forward, and it's doubtful that environmentalists would be pleased with very many of them," says Michael Livermore, executive director of the Institute for Policy Integrity at the New York University School of Law. "It's bare politics. They want to enact as much of their agenda [as they can] before they get out of office."

Although incoming presidents can in some cases block 11th-hour regulations, it can be difficult for agencies to reverse course in a rule-making process that is, theoretically, separate from the political fray.

When Bush came to office in 2001, he immediately put a hold on all regulations
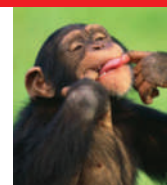
Pfizer's decision to eliminate cholesterol research comes as Lipitor, one of a class of cholesterol-reducing drugs known as statins, brought in US$12.7 billion last year as the world's best-selling drug. Still, "the cholesterol market is totally saturated," says Jason Bowers, an analyst for the market-research firm Decision Resources in Waltham, Massachusetts. "And as soon as Lipitor goes off patent, the statin market is going to fall."

Industry in general is pulling away from research on cardiovascular disease, even though it remains the leading cause of death in the United States. A recent analysis of clinical trials from 2005 to 2007 revealed a decline in the number of drug trials against leading cardiovascular conditions such as high blood pressure and high cholesterol (J. P. E. Karlberg, *Nature Rev. Drug Discov.* **7**, 639–640; 2008).

That's not totally surprising, says Damien Conover, an analyst for Morningstar, an investment-research company based in Chicago, Illinois. Recent scandals about the dangerous side effects of drugs, such as the increased risk for heart attack and stroke in those taking Merck's painkiller Vioxx, have made the US Food and Drug Administration more cautious about drug approvals. Drugs that have been on the market for years are more familiar and often viewed as less risky, meaning that

DRUG INDUSTRY RIDES THE STOCK MARKET
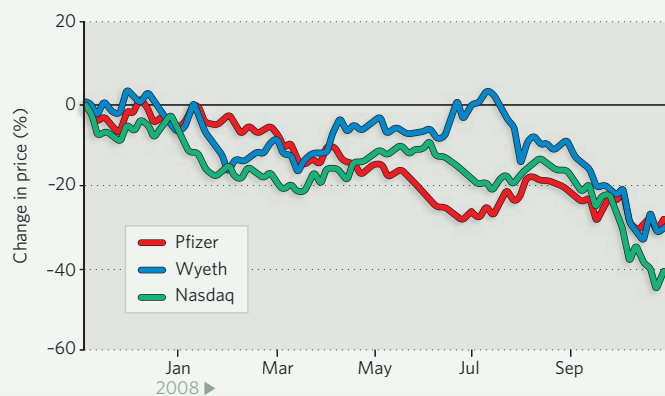
new drugs must perform much better than those already on the market before regulators are willing to take a chance on approval. Meanwhile, health-care insurance plans also view new medications with increased scrutiny. In crowded fields such as cardiovascular disease and other primary-care areas, the hurdle is simply too high for companies to risk it.

This makes specialized fields such as immunology and neurology attractive alternatives. The number of cases of rheumatoid arthritis, for example, has increased faster than expected, generating a large population of patients that companies are now scrambling to target.

And some predict that the next blockbuster drug will be a treatment for Alzheimer's: a disease with a growing market, a tremendous unmet need for treatments, and a cadre of patients and family members willing to pay

high prices for therapy. Wyeth has been a leader in Alzheimer's research, both in therapeutics and in development of a candidate vaccine. "People think that in the next five or eight years, there's likely to be a major advance, and if you're the first company with that advance, the payout will be gigantic," says Erik Gordon, associate dean of technology management at the Stevens Institute of Technology in New Jersey.

Meanwhile, research investment in oncology has been growing steadily across the industry. Some view Indiana-based Eli Lilly's recent $6.5-billion bid for the biotechnology firm ImClone as a sign of increased demand for cancer drug candidates. Health-care insurance plans, too, have traditionally been more willing to pay high premiums for cancer therapies — although there are signs that this attitude may be changing. And the pharmaceutical industry has recently embraced the drive towards genetically targeted, individual treatments in oncology — a concept that once made companies cringe because it reduced the market for a given drug.

That, says Conover, was before the industry realized that patients would pay tens of thousands of dollars for an expensive new drug. "All of a sudden," he says, "'market limiting' is OK." ■

Heidi Ledford

---

passed shortly before Bill Clinton left the White House. Even so, some of Clinton's policies endure: a rule protecting almost 60 million acres of roadless areas, for instance, remains in force even though the legal battle continues eight years later.

This year, White House chief of staff Joshua Bolten issued a memo on 9 May directing all agencies to propose any new regulations before 1 June. But the administration proposed several controversial regulations after that date. One of these would scale back the requirement for Endangered Species Act consultations with federal biologists on projects such as roads and pipelines. Another would require the Department of Labor to conduct risk assessments for toxic chemicals on an industry-by-industry basis. The new rule would make such assessments more difficult, Livermore says, and the

US mining regulations could become less stringent.

resulting standards less protective.

In trying to keep the last-minute rule-making to a minimum, Bolten also directed that regulations be finalized by 1 November. Many are, however, still working their way through the system. "What we are learning is that the deadlines are not very firm," says

Rick Melberth, who heads federal regulatory policy at OMB Watch.

Agencies can sometimes pull back rules that have not come into effect, generally within 30 to 60 days of their being issued. But once a rule has come into effect, Melberth says, the administration must start over — or simply dedicate fewer resources to its implementation.

White House officials say that Bolten's memo was never intended as a moratorium on regulatory activity. Many of the rules have been under discussion for years and could be finalized in the next administration. "It's a matter of due diligence," says Jane Lee, spokeswoman for the Office of Management and Budget, which oversees federal regulatory changes. "We are going to make sure that all regulations have the benefit of a thorough and full review." ■

Jeff Tollefson

NASA/JPL-CALTECH/UNIV. ARIZONA/TEXAS A&M UNIV.

PHOENIX FADES AWAY

Summer's lease is ended:
Phoenix will dig for ice no more.

NASA/JPL-CALTECH/CORNELL; NASA; NASA/JPL-CALTECH/UNIV. ARIZONA/MSSS

NASA/JPL-CALTECH/UNIV. ARIZONA/MAX PLANCK INSTITUTE

TUCSON

In its final days Phoenix, the NASA lander that since May has been scraping at subsurface ice in the martian arctic, is blinking in and out of contact with Earth.

As temperatures plummeted to nearly −100 °C and dust storms and clouds obscured an enfeebled sun, the spacecraft last week missed several chances to communicate with satellites passing overhead, and plunged for the first time into a bare-bones survival mode. Although engineers may wring a few more days of erratic behaviour out of the lander, Phoenix is almost certain not to survive the coming winter. Thick slabs of frozen carbon dioxide will coat the spacecraft, and its electronics will break for good. "It's like an ageing parent in the nursing home. You know it's coming," says principal

investigator Peter Smith of the University of Arizona in Tucson.

Engineers hope to initiate a final sequence of low-power experiments — mostly weather measurements — as early as 5 November, but the craft's weak condition may not even allow this. With a sudden surfeit of free time, mission scientists are returning to their notebooks to see what their data may say about the history of martian water ice and its implications for habitability. But frustrations with some of the instruments mean that the story is not complete.

For Smith, the highlight was Phoenix's flawless landing in May — a redemption for a spacecraft that was mothballed in a Lockheed Martin warehouse after a sister ship, the Mars Polar Lander, crashed near the south pole of Mars in 1999. The bold idea for Phoenix

was to free the spacecraft from storage, strip it of Polar Lander's bugs and rebuild it within a $420-million budget. Operated from the University of Arizona, the lander would also be the first major mission controlled from outside of NASA's traditional centres such as the Jet Propulsion Laboratory in Pasadena, California. "There were a lot of doubters," says Smith. "It wasn't a sure thing by any means."

Yet Phoenix settled successfully on a northern plain. Originally scheduled for three months, the mission was then extended to at least 18 November, and the early part went swimmingly. A robotic arm and scoop clawed its way through a few centimetres of soil to water ice. A camera, perched atop a tall mast, caught nuggets of ice sublimating away. Panoramas showed differently sized polygonal cracks in the soil, suggest-

THE PHOENIX MARS LANDER: A LIFE

4 August 2007
Launches atop a Delta rocket from Cape Canaveral, Florida.

25 May 2008
Lands at 68° north on Mars.

1 June
Scoops its first martian soil.

30 July
Confirms that white substance in soil is water ice.

5 August
Identifies the oxidizing compound perchlorate in the soil.

LARGE HADRON COLLIDER
How to repair the biggest science experiment in the world.
www.nature.com/news

ing that contractions due to freezing occurred over different temperature regimes in recent epochs, as the tilt and orbit of Mars changed. And a wet-chemistry instrument, using beakers that made soil slurries with water brought from Earth, found that the polar soil was like nothing else tested so far on Mars. The soil was basic, rather than acidic, and contained trace amounts of perchlorate, a weak oxidizer that on Earth can nourish microbes.

But Phoenix's workhorse instrument, a unit with eight ovens that baked thimbles of soil and could sniff emitted gases for organic compounds, was plagued with problems. The soil, surprisingly sticky, was hard to get past oven doors that, owing to a manufacturing error, opened only partway. A short circuit, probably caused by shaking the ovens to move the sticky soil, made NASA nervous. Worried that the instrument could fail at any time, headquarters directed the team to prioritize retrieving an ice-rich sample for analysis.

"The clock was running against us," says William Boynton of the University of Arizona, lead scientist for the instrument. Trying to nab the ice sample was particularly frustrating. "We wasted nearly half of the mission doing that," he says. In the end, Boynton got results from only five of the eight ovens. He never got to test the ice for isotopic ratios that could have said something about its age.

However, towards the end of the mission the instrument began to redeem itself. It found a strong signal for calcium carbonate, a mineral that typically forms in the presence of water. A separate, weaker signal in the soil may indicate a different type of carbonate, or even an organic molecule. Boynton hopes to settle this by comparing carbon isotopes from the two sources. With the stress of daily operations halting, he'll finally have a chance to do so. ■

Eric Hand
For an online slideshow of pictures taken by Phoenix, see http://tinyurl.com/6a3vkt.

Human genes are multitaskers

Although people often struggle to master more than one discipline, our genes are accomplished polymaths. Genome-wide surveys of gene expression in 15 different tissues and cell lines have revealed that up to 94% of human genes generate more than one product.

The surveys, published online on 2 November in *Nature*[1] and *Nature Genetics*[2], used high-throughput sequencing to generate the most detailed portrait yet of how genes are expressed in different tissues.

Only about 6% of human genes are made from a single, linear piece of DNA. Most genes are made from sections of DNA found at different locations along a strand. The data encoded in these fragments are joined together into a functional messenger RNA (mRNA) molecule that can be used as a template to generate proteins.

But researchers have found that the same gene can be assembled in different ways, sometimes leaving out a piece, for example, or including a bit of the intervening DNA sequence.

This process, called alternative splicing, can produce mRNA molecules and proteins with dramatically different functions, despite being formed from the same gene. The phenomenon provides some solace to those disappointed by the relatively small number of genes in the human genome: with around 20,000 genes, humans have roughly the same number as the elegant but decidedly less complex nematode, *Caenorhabditis elegans*.

"We were expecting that something as sophisticated, complex and intelligent as ourselves would have about a hundred thousand genes at least," says Jacek Majewski, a genomicist at McGill University in Montreal, Canada. "Then we sequenced the genome and realized it was about the same number as *C. elegans*." Fortunately, alternative splicing is thought to occur in only about a tenth of *C. elegans* genes, restoring the dignity of complexity to the human genome. Understanding this flexibility should help to reveal how improperly spliced genes can trigger disease.

Despite intense interest in alternative splicing, the phenomenon has been difficult to study, and the usual laboratory techniques often fail to detect rare splice forms.

Researchers previously estimated that 74% of all human genes are alternatively spliced[3], but recognized that this estimate was likely to increase as techniques to study the process improved.

Now two groups, one led by computational biologist Christopher Burge of the Massachusetts Institute of Technology in Cambridge and the other led by molecular biologist Benjamin Blencowe of the University of Toronto in Canada, have studied alternative splicing using high-throughput sequencing data generated by Illumina, a biotechnology company based in San Diego, California.

The technique works by using an enzyme to convert mRNA back to DNA, which can then be sequenced. Blencowe and his colleagues studied splice forms found in six different tissues, including the brain, liver, muscle and lungs. Burge and his colleagues used these samples along with several others, including breast cancer cell lines. On the basis of more than 400 million sequences, Burge's team estimates that 92–94% of all human genes can yield more than one RNA molecule.

Specialists in the field agree that the work is important, but are not particularly surprised by the numbers. "What is new is the technology, which will have a big effect on how we study splicing," says Douglas Black, a molecular biologist at the University of California, Los Angeles.

Analysis of the new splicing catalogues can reveal patterns about how the process is regulated, but more work is needed to determine whether all these splice forms have a function. "The question now is, 'Are all of those forms biologically relevant?'" says Marie-Laure Yaspo, a genomicist at the Max Planck Institute for Molecular Genetics in Berlin, Germany. A few of those rare splice variants may be no more than background noise generated by occasional mistakes, she notes.

But conventional techniques for deleting entire genes are not effective for sorting out the function of one splice variant from another. "What really needs to be done is to develop high-throughput methods for analysing the function of these splice variants," says Blencowe. "That's the big challenge ahead." ■

Heidi Ledford

1. Wang, E. T. *et al. Nature* doi:10.1038/nature07509 (2008).
2. Pan, Q. *et al. Nature Genet.* doi:10.1038/ng.259 (2008).
3. Johnson, J. M. *et al. Science* **302**, 2141–2144 (2003).

Christopher Burge.

29 September
Detects evidence for calcium carbonate, a mineral that forms in liquid water.

28 October
Communications become erratic and science operations cease, possibly for good.

8 September
Photographs dust devils in the atmosphere.

How to get the most from a gene test

According to two commercial gene-testing services — 23andMe and deCODEme — US Army medic Timothy Richard Gall of Fort Belvoir, Virginia, has a higher-than-average risk of basal cell carcinoma, type 2 diabetes and psoriasis. But much more enlightening than these results, which cost Gall more than $1,400, was a free online program called Promethease that he used to further analyse the data. By offering more in-depth information and interpreting of more of his genetic variants, Promethease "gives a much more realistic view of the usefulness of the information", Gall says.

Start-ups and services such as Promethease are now developing ways to improve the limited value of information provided by personal genomics companies for consumers and scientists alike.

For instance, Omicia, based in Emeryville, California, is designing software to make sense of entire genome sequences, such as those of the individuals published in this issue (see pages 53 and 60). At present, firms offering genetic testing look only at small variations called single nucleotide polymorphisms, or SNPs. But people looking at their whole genomes will also want to know the meaning of all the different types of variation within them, such as extra copies of genes or flipped sections of DNA. Omicia examines each location in a person's genome and compares it to the company's own analysis of disease risks linked to all the types of variation known to exist. "We've always had the full genome in mind, so for us, any kind of position somebody finds we can link to their genome and to our system," says company co-founder Martin Reese.

Of course, most consumers are still stuck with SNPs, and Promethease attempts to squeeze as much information as possible out of these. The program uses data compiled in a wiki called SNPedia, launched in 2006 and run in the spare time of bioinformatician Mike Cariaso and scientist/entrepeneur Greg Lennon. Their wiki contains information selected from the vast public databases commonly used for research, such as dbSNP, and tries to make it more useful by, for instance, including short

23andMe analyses saliva collected in home test kits (above) for small genetic variations.

Mike Cariaso runs an online program that analyses commercial gene-test results.

written interpretations of the SNPs' importance in various health conditions on the basis of published studies. Compared with reports delivered by gene-testing companies, Promethease reports are more detailed and nuanced — containing information on, for example, more SNPs, how common each of a person's particular genetic variants are and the magnitude of the likely impact of each variant.

Cariaso, who lives in Bethesda, Maryland, says that the ability to link genes to traits through SNPedia will become more useful with more individuals' data — so he has begun analysing data from the Personal Genome Project (PGP). This aims to sequence and post the genomes of as many people as possible, along with data about their medical, mental and physical characteristics. PGP released its first batch of data on 21 October, to grumblings about its quality. Two days later, Daniel MacArthur, a postdoc at the Wellcome Trust Sanger Institute in Cambridge, UK, wrote in his blog that the data were "pretty underwhelming", containing mostly low-quality sequence information on just four people that covers only 0.13% of the entire genome. "Given

the hype surrounding this data release I'm a little disappointed by the data itself," he wrote.

George Church, a Harvard University molecular geneticist who runs the PGP, doesn't disagree. "You should be underwhelmed" by the first data release, he says. He calls it "really more of a social engineering event than a true production announcement". Right now, Church says, the main focus is recruiting more study participants to improve the project's scope and its data quality; he adds that 9,500 people have now volunteered.

Already, Church's 'social engineering event' has stirred a public dialogue about the usefulness of linked genetic and medical data. For instance, Gall is one of seven people who have released their Promethease reports publicly on SNPedia independently of any research project.

Gall says he posted his data in part because he knows its value today is still limited, and he wants that to change: "By making it public, I hope that I will only increase its usefulness to me personally," Gall says. That shouldn't be too hard: Gall notes that his SNPs didn't even reveal the makeup of some medically crucial genes — such as those that determine his compatibility for organ donations — whose composition he can learn for free as a member of the military: "In that sense, 23andMe and deCODEme are not worth the price," he says. ∎
Erika Check Hayden

Q&A

Genomics takes hold in Asia

Recruited in 2001 from the US National Cancer Institute, **Edison Liu** was the first big international catch for Singapore's burgeoning Biopolis research hub. He still heads the Genome Institute of Singapore there and had the leading role in the Pan-Asian SNP Initiative, an effort to compare subtle genetic variations across Asian populations. He is currently chairman of Singapore's Health Sciences Authority and president of the Human Genome Organisation (HUGO). He spoke with *Nature*'s David Cyranoski about how to make pan-Asian genomics research projects work.

Collaborations among Asian scientists are just not as strong as those they share with scientists in the West. Why?
Scientists in Asia have a tendency to look past each other and focus on collaborations with the United States or Europe, partly because these collaborations get them more credit from their school administrations. Also, in Asia, most countries see each other as competitors. Just getting people together is an accomplishment.

About seven years ago we began talking about doing a genetics project in the Asia–Pacific. It didn't gel for several reasons. The science itself wasn't mature. There was also a great disparity in capabilities and access to technologies between the various countries. There were also ethical concerns; some indigenous populations were worried about being told they were more likely to have certain diseases, and what that might mean for individuals within those populations.

How did you get around that?
We removed the negative connotations of disease and phenotype and just focused on how genetically different people were. Instead of looking at disease, we just found something that Asian scientists can work on together. Once we focused on diversity, all of a sudden countries such as Indonesia and the Philippines became very important because of the inherent diversity in their national populations. The infrastructure advantages of some countries such as Japan and Korea was evened out by the diversity that others offered.

What were the results?
We had data from 11 countries, 30 institutions and more than 73 ethnic groups. It was an interesting repainting of the migration history of Asia using autosomal markers. It confirmed some hypotheses, for example that genetic clustering was correlated with linguistic clusters. It also raised some questions. For example, our analyses suggested a common ancestral origin for all East and Southeast Asian populations studied, implying that most of the gene pool in Asia came from a single initial entry of modern humans into the continent. The paper has now been submitted for publication.

Partly as a result of this, some countries have become more active in genetics. The Philippines health ministry is contemplating putting more resources into its genetics institute arm. The Eijkman Institute in Jakarta has begun to get more international funding. Twenty years ago, the only place recognized for genetics in Asia was Japan.

After the heyday of 2003 when the complete human genome was announced, HUGO dropped off the map. You have now moved its official base to Singapore. What do you hope to achieve?
HUGO was ready for a re-look at its mission and goals. Much of the move of the HUGO office to Singapore was because of practical matters — for ease of administration. However, part of the move was a recognition that the strongest chapter of HUGO was the Asia–Pacific. The core of scientific activity is more distributed now and is shifting eastwards.

There is also already talk of a phase II for our international SNP [single nucleotide polymorphism] initiative. For the current project we used 50,000-SNP microarrays from Affymetrix, which covered autosomal markers. Going forwards, we would like to use 500,000- or 1-million-SNP arrays. We could look at copy-number variation and haplotypes in the populations.

Now the first Asian genome has been sequenced (see page 60). The novelty and the timeliness of the catalogue of information makes it worthwhile. This is just the beginning of global efforts to sequence more human genomes so that we understand the range of diversity in our species. In the future, I would like to pursue a Pan-Asian genome project. Knowledge of this diversity is important as we try to match the best therapeutic drugs to specific world populations.

What's next?
I'd like to extend such a diversity study to Central Asia and across the Pacific Rim to some American Indian populations. With this breadth of coverage, the scientific community could really get a sense of the variability of the genome and how it has changed with migration. We could also start getting a genetic clock to go along with the anthropological ones that we have in mapping the history of humanity.

What value will this have for biomedicine?
Of course this will eventually help with clinical trials. We don't have the phenotypic data now, but as association studies get richer, certain genes will be linked to a drug response or drug toxicities. It is conceivable that with a map of human genetic diversity, health planners can project whether certain therapeutic drugs will be more or less appropriate for specific populations. ■
See Editorial, page 1.

R. NICKELSBERG/GETTY IMAGES

Legally binding green targets for UK

The United Kingdom is on track this month to become the first country to have legally binding targets for cutting greenhouse gases.

On 28 October, the British Parliament approved a bill that includes a requirement to cut carbon dioxide emissions to 80% below 1990 levels by 2050. The bill is expected to receive royal approval this month, making it law.

"This has not been done anywhere in the world," says Jim Skea, research director of the UK Energy Research Centre in London.

Members of Parliament voted in favour of the bill after the government amended the legislation to include emissions from the aviation and shipping industries within five years. The bill sets out a broad framework within which Britain will draw up specific plans to reduce its carbon footprint.

The legally binding limits on carbon emissions will be set at five-year intervals, although it is not yet clear what penalties missing the targets would incur. And the government will have powers to set up carbon-trading schemes, to encourage firms to reduce pollution.

The bill creates an independent advisory committee on climate change made up of leading scientists and economists and led by Adair Turner, a businessman and member of the House of Lords. The group will suggest interim emissions-reduction targets leading up to 2020; recommend the levels at which the five-year limits should be set; and suggest actions needed by different sectors of the economy.

> **"The interim target decided for 2020 will have a really big impact on the business community now."**

It will report next month to the government, which will decide by March 2009 whether to accept the suggestions.

Skea, a member of the new committee, says the bill's immediate effect will depend heavily on the interim target decided for 2020. That, he says, "will have a really big impact on the business community now".

Environmental activists praised the bill, but noted that the government is still planning to build a coal-fired power plant in Kingsnorth in southern England — a move that has met with strong protests.

Meanwhile, the European Union (EU) is pushing ahead with its own plans to modify the EU emissions-trading scheme, which is Europe's key mechanism for reducing greenhouse gases. On 24 October, EU ministers formally agreed that aviation will be included in the scheme, starting in 2012. Aircraft produce about 3% of Europe's greenhouse gases, and aviation emissions have increased by 87% since 1990.

"Bringing airlines into the EU emissions-trading scheme will provide a real incentive for airlines to reduce their carbon emissions," says Ed Miliband, the UK energy and climate-change secretary.

The British Air Transport Association, an industry group, called the UK bill disappointing in focusing on national aviation limits. Roger Wiltshire, the group's secretary-general, said in a statement that the Europe-wide initiative instead represented a "sensible approach". ■
Natasha Gilbert

Australian government plans Internet censorship

The government in Australia is proposing to introduce compulsory blocks on certain websites for everyone accessing the Internet from inside the country, and to trial these filters before the end of the year.

Stephen Conroy, the communications minister, said in a Senate hearing in late October that the government proposes a two-tier system of restrictions as part of its Aus$44.2 million (US$29.8 million) cyber-safety plan. The first tier of filters would be compulsory, and would force all Internet service providers to block Australians' access to illegal websites, including overseas online gambling sites, according to a report in the *Sydney Morning Herald*. A second tier is planned as an optional set of additional filters that would make it impossible to access material deemed unsuitable for children.

Religious organizations such as the Australian Christian Lobby support the proposals. But critics worry about the ease with which the government might exand the list of blocked sites. They claim that access will not be effectively cut off to illegal content such as child pornography because peer-to-peer file-sharing networks will remain unimpeded. The System Administrators Guild of Australia says the planned filters will increase the price and reduce the speed of Internet access.

Oil company blamed for mud-volcano eruption

After more than two years of debate, a vote of 74 Earth scientists at last week's American Association of Petroleum Geologists conference blamed an Indonesian oil company for creating a mud volcano.

Mud has been spewing from a former rice paddy in Sidoarjo in East Java since 29 May

The Sidoarjo mud volcano has caused widespread devastation.

Repair puts Hubble back on track

The Hubble Space Telescope's system for storing and transmitting data, which failed in September, has been restored to service. Among the first new images released is this pair of interacting galaxies in the constellation Cetus, more than 100 million parsecs from Earth.

A servicing mission to the telescope has been delayed from February 2009 until at least May.

NASA/ESA/M. LIVIO/STSCI

2006 (see *Nature* 445, 812–815; 2007), and the question of who will pay for the clean-up hangs on what caused the disaster. PT Lapindo Brantas, the oil company, says the cause was an earthquake that had struck two days beforehand.

But the majority of scientists attending the meeting in Cape Town, South Africa, voted that the tremors had hit too far away for them to be responsible. Some researchers presented data showing that the pressure created by the company's drilling was sufficient to break a path for deep mud to rupture the surface.

Help promised for troubled maker of electric car

Elon Musk, chief executive of electric-car company Tesla Motors, has denied rumours that the company is about to go bust.

In an interview on 31 October with Reuters news agency, Musk admitted that Tesla Motors, which has just started making its environmentally friendly sports cars, has only US$9 million in the bank, despite taking deposits on 1,200 roadsters costing $109,000 each. The interview was sparked by comments posted on the Valleywag blog, a Silicon Valley gossip site, suggesting that Tesla would not be able to deliver the promised cars.

On 2 November, Tesla, which is based in San Carlos, California, announced that several unnamed investors have promised $40 million to help prop up the firm.

Google settles suit over copyright of scanned books

Google has agreed to pay US$125 million to settle a class action suit brought by the Authors Guild and several publishing companies in the United States against its Google Book Search service. Authors and

publishers will also get a cut of the future revenues that the service generates.

In return, the 28 October deal permits Google's US readers to see fuller previews instead of small snippets of copyrighted books. Libraries will be able to subscribe, which will allow their patrons to read the contents of entire books over the Internet. Text and data-mining researchers will also have the right to run computational queries in a 'research corpus' copy of the entire Google Book text and image database.

Google has already scanned into the database more than 7 million titles from university research collections and partner publishers.

Crop research a target of international investment

The US National Science Foundation (NSF) has awarded nearly US$60 million in grants for plant-genome science, most of which will go towards research in crop species.

On 27 October, the NSF assigned $3.2 million for work probing how genetic and biochemical pathways help *Medicago truncatula* — a model organism for legume research — adapt to high-salinity conditions. That project is being led by scientists at the University of Southern California. Another of the 20 grants was given to researchers at Pennsylvania State University in University Park. The team received $4.8 million to look into the genes that function during the growth of maize (corn) shoots, which is regulated by the hormone auxin.

And in Britain, the government's science think tank, the Foresight Programme, wants to know how to feed nine billion people equitably, healthily and sustainably. It is launching a study into what the world's farming needs might be in 2050, the findings for which should be available in 2010.

S. PAMUNGKAS/REUTERS

America's new leadership

Researchers should keep a cool head about science under Obama, **David Goldston** argues.

Editor's note: This column, which went to press before the 4 November US election, is written based on the poll indicators at the time — which pointed strongly towards a Democratic victory.

The US presidential campaign goes on for so long that it's hard to imagine life without it. But pundits and policy advocates can adjust quickly, shifting from speculating about who the president will be to speculating about what he will do And when it comes to science, the victory of Barack Obama and the Democrats in Congress has created an enormous sense of anticipation.

Scientists' most immediate, and perhaps most fervent, concern is funding, and Obama's election does indeed portend better times. But for energy research, it is likely to be years before the numbers match the campaign rhetoric. Obama's proposal was to spend US$15 billion a year on new energy technologies, research that receives only about $2 billion a year today. Less advertised was that the new money would be raised through climate-change legislation. Under Obama's plan, the government would auction permits to allow industry to emit greenhouse gases under a bill to cap emissions.

But climate-change legislation remains controversial even among Democrats, and the economic downturn will make it even harder to enact early in the new administration. Moreover, Congress will probably allocate at least a portion of the permits for free, reducing the funds Obama wants to tap for research. And support is growing among economists and politicians for the idea of rebating most, if not all, of any auction revenue to taxpayers to alleviate the impact of higher energy prices, rather than using the money for government programmes.

Delaying a massive spike in energy spending might not be an entirely bad thing. The details of energy research programmes haven't been seriously rethought in years, and experts disagree on where government funds would be most helpful. In Congress, the discussion has largely been stuck on the philosophical question of whether government research can advance useful technology or does damage by distorting market forces. That debate has not progressed much since Ronald Reagan tried to eliminate most applied energy research

PARTY OF ONE

in the 1980s. Allowing time for a thoughtful review before spending $15 billion — an arbitrary figure, in any event — couldn't hurt. In the meantime, there is plenty of energy research that could benefit now from a more incremental influx of cash, including work on batteries, commercial building efficiency and carbon sequestration.

Money is likely to be available for such initiatives. The financial crisis and economic slowdown will probably contribute to a boost in research spending. Concerns about the ballooning deficit are being eclipsed by the push to use government spending to stimulate the economy. And the size of the total domestic spending pie — which Obama wanted to enlarge even before the Wall Street meltdown — is always the best indicator of how much will be allocated to science.

Beyond that, science advocates will no doubt contend that research spending should be especially favoured in any economic stimulus package because it contributes to future economic growth. That line of argument may get science still more money even though research doesn't fit the profile of ideal stimulus spending — programmes that quickly get money into the hands of lower- and middle-income consumers who will spend it most rapidly.

So the question doesn't seem to be whether research budgets will fare better under Obama, but rather by how much. The budgets of the National Science Foundation, the National Institute of Standards and Technology (NIST) and the Office of Science at the Department of Energy are likely to be put on a path to double over 10 years, a move that both President George W. Bush and the Democratic

Congress have supported in principle. And doubling spending at the three agencies is relatively cheap; together they now spend about $11 billion a year.

Obama has also called for a 10-year doubling of the budget of the National Institutes of Health (NIH), now about $30 billion a year. The NIH is a popular cause, but it could face tough competition for dollars even in an expansive climate. The agency is funded through the same spending legislation that finances education and social-services programmes, which have more pent-up demands, larger constituencies and a higher profile. And the highest Obama priorities in health are improving care and expanding insurance coverage, not research. Also, in the wake of the doubling of the NIH budget from 1998 to 2003, Congress has begun to wonder aloud whether the additional spending resulted in enough tangible benefits for patients.

Increases for other agencies pose their own conundrums. For example, pushed by the need to garner votes in Florida and concerns about relying on Russia for its Soyuz spacecraft in the wake of the invasion of Georgia, Obama called for prolonging the life of the space shuttle and reducing the gap between the shuttle's retirement and the launch of a new vehicle for sending humans to the Moon. Doing that while launching more scientific satellites, which Obama also supports, would require a significant increase in NASA's $17-billion budget. Congress could baulk at such a boost, leaving the agency in its usual predicament of too many missions and too little cash.

But the biggest unknown about science under Obama is what new initiatives he will propose. Unlike Bush, Obama has no philosophical qualms about government programmes to stimulate industrial innovation. Existing efforts such as the Technology Innovation Program (formerly the Advanced Technology Program) at NIST will probably get a new lease of life, and proposals from think tanks, such as the creation of a national innovation foundation, could get a closer look.

So the air of anticipation in the nation's laboratories and faculty clubs is not unfounded; the danger is that it will become excessive. Like all presidents, Obama will have to govern as a mere mortal, making trade-offs among legitimate claims on the public purse and crafting political deals among constituencies. Scientists are going to have to tame their insatiable appetite for dollars, and their tendency to see politicians as either with them or against them, for the current mood to survive much beyond the inauguration. ∎

David Goldston is a project director with the Bipartisan Policy Center in Washington DC.

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have estimated that height is 80–90% heritable. So if 29 centimetres separate the tallest 5% of a population from the shortest, then genetics would account for as many as 27 of them[1].

This year, three groups of researchers[2–4] scoured the genomes of huge populations (the largest study[4] looked at more than 30,000 people) for genetic variants associated with height differences. More than 40 turned up.

But there was a problem: the variants had tiny effects. Altogether, they accounted for little more than 5% of height's heritability — just 6 centimetres by the calculations above.

Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did "very little of the prediction that you would do just by asking people how tall their parents are", says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies[3].

Height isn't the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism[5] and more than 80% for schizophrenia[6]. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see 'Where's the reward?', page 20). And to some extent they have, identifying more than 400 genetic variants that contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. "It is the big topic in the genetics of common disease right now," says Francis Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point "where we all had to scratch our heads and say, 'Huh?'", he says.

Although flummoxed by this missing heritability, geneticists remain optimistic that they can find more of it. "These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability," says Hirschhorn. So where might it be hiding?
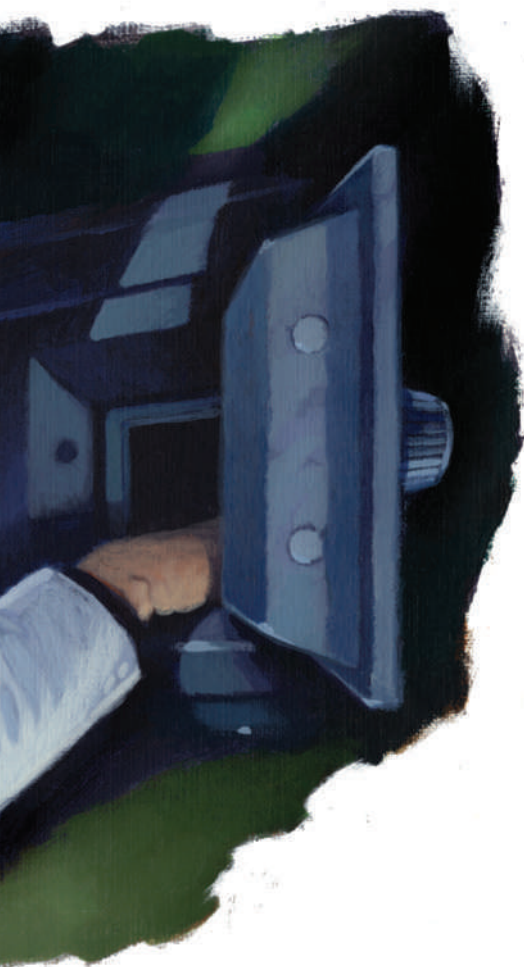
Right under everyone's noses

The inability to find some genes could be explained by the limitations of GWAS. These studies have identified numerous one-letter variations in DNA called single nucleotide polymorphisms (SNPs) that co-occur with a disease or other trait in thousands of people. But a given SNP represents a much bigger block of genetic material. So, for example, if two people share one of these variants at a key location, both may be scored as having the same version of any height-related gene in that area, even though one person actually has a relatively rare mutation that has a huge effect on height. The association study might identify a variant responsible for the height difference, says Teri Manolio, director of the Office of Population Genomics at the NHGRI, but averaging across hundreds of people could give the appearance that its effects are pretty wimpy. "It's going to be diluted," she says.

Finding this type of missing heritability is conceptually easy, because it involves closer scrutiny of the genes already in hand. "Just exploring, in a very dense way, genetic variation at the loci that have been discovered is probably going to [explain] another increment of missing heritability," Hirschhorn says.

Researchers will need to sequence candidate genes and their surrounding regions in thousands of people if they are to unearth more associations with the disease.

Helen Hobbs and Jonathan Cohen of the University of Texas Southwestern Medical Center in Dallas did this in an attempt to capture all the variation in *ANGPTL4*, a gene their studies had linked to cholesterol and triglyceride concentrations. They sequenced the gene in around 3,500 individuals from the Dallas Heart Study and found that some previously unknown variants had dramatic effects on the concentration of these lipids in the blood[7]. Mark McCarthy of Britain's Oxford Centre for Diabetes, Endocrinology and Metabolism says that such studies could reveal much of the missing heritability, but not a lot of people have had the enthusiasm to do them. This could change as the cost of sequencing falls.

Out of sight

Other variants, for which GWAS haven't even begun to provide clues, will prove even harder to find. In the past, conventional genetic studies for inherited diseases such as cystic fibrosis identified rare, mutated genes that have a high penetrance, meaning that the gene has an effect in almost everyone who carries it. But it quickly became apparent that high-penetrance variants would not underlie most common diseases because evolution largely keeps them in check.

What powered the push into genome-wide association was a hypothesis that common diseases would be caused by common, low-penetrance variants when enough of them showed up in the same unlucky person. Now that hypothesis is being questioned. "A lot of people are recognizing that screening for common variation has delivered less than we had hoped," says David Goldstein, professor of genetics at Duke University in Durham, North Carolina.

But between those variants that stick out like a sore thumb, and those common enough to be dredged up by the wide net of GWAS, there is a potential middle ground of variants that are moderately penetrant but are rare enough that they are missed by the net. There's also the possibility that there are many more-frequent variants that have such a low pen-etrance that GWAS can't statistically link them to a disease.

These very-low-penetrance variants pose some problems, says Leonid Kruglyak professor of ecology and evolutionary biology at Princeton University in New Jersey. "You're talking about thousands of variants that you would have to invoke to get near 80% or 90% heritability." Taken to the extreme, practically every gene in the genome could have a variant that affects height, for example. "You don't like to think about models like that," Kruglyak says.

If rare, moderately penetrant or common, weakly penetrant variants are the culprits, then bumping up the number of people in existing association studies could help find previously missed genetic associations. Peter Visscher of the Queensland Institute of Medical Research in Brisbane, Australia, says that a meta-analysis of height studies covering roughly 100,000 people is in the works. Lowering the stringency with which an association is made could drag up more, but confidence in the hits would drop.

At some point it might make sense to stop using SNPs, and start sequencing whole genomes. Collins suggests that the NHGRI's 1,000 genomes project, which aims to sequence the genomes of at least 1,000 people from all over the world, could go a long way towards finding hidden heritability, and many more genomes may become possible as the price of sequencing falls.

Not everyone supports an all-out sequencing onslaught. Goldstein warns against

Where's the reward?

There is more riding on the case of missing heritability than academic satisfaction. By finding variants related to common disease, genome-wide association studies promised to deliver meaningful medical information and justify the US$3 billion spent on the human genome and the multimillion-dollar effort to map human variation. "The reason for spending so much money was that the bulk of the heritability would be discovered," says Joseph Nadeau, a geneticist at Case Western Reserve University in Cleveland, Ohio.

The ability to predict someone's height from their genes would be a pretty trivial carnival trick, but it represents a mastery over the language of life that could potentially spill into most areas of medicine. Aside from some surprises, though, such as mutations in immune-system genes being tied to an eye disorder called age-related macular degeneration, many of the variants found have only modest effects on human characteristics. For now, genetics rarely provides a clearer predictive answer than a good family history. And the path to therapy is not straightforward, says David Goldstein of Duke University in Durham, North Carolina. "This talk about personalized risk profiles, using genetics, for most common diseases, and this talk about a whole flood of new drug targets. I think that that's now pretty clearly wishful thinking."

Francis Collins, former head of the National Human Genome Research Institute in Bethesda, Maryland, agrees that the picture for disease prediction remains bleak, but is still optimistic about therapeutic intervention. Most genetic variants found by genome-wide association "contribute a relatively modest risk, but that in no way says the genes aren't important," he says. "The opportunity for therapy here is breathtaking."

Peter Visscher, a geneticist at the Queensland Institute of Medical Research in Brisbane, Australia, agrees. "It would be easy to knock [genome-wide association studies] and say everything was promised and nothing was delivered. But in terms of identifying genes and pathways for disease, it's been very successful. I would feel it's moved the field forwards tremendously."

Ultimately, the clinical value of the variants that genome-wide association studies have turned up may differ from disease to disease. Still, some say that the field is too fixated on clinical application, be it through prediction, personalization or identifying drug targets. Robert Nussbaum of the University of California, San Francisco, puts it bluntly: "Human genetics research always assumes too quickly that it has to be translational. They're doing basic research." **B.M.**

---

continuing to "turn the crank" without devising a more rational approach, such as sequencing the genomes of people who exhibit extreme manifestations of diseases. "I'm not really sold on doing the sequencing version of what we did with [GWAS]," he says. "It's a big enough, costly enough job, that I think we want to think a little bit harder about exactly who gets re-sequenced."

In the architecture
Some researchers are now homing in on copy-number variations (CNVs), stretches of DNA tens or hundreds of base pairs long that are deleted or duplicated between individuals. Variations in these features could begin to explain missing heritability in disorders such as schizophrenia and autism, for which GWAS have turned up almost nothing. Two recent studies looked at hundreds of CNVs in normal people and in those with schizophrenia, and found strong associations between the disease and several CNVs[8,9]. They commonly arise *de novo* — in an individual without any family history of the mutation.

These structural variants might account for a lot of the genetic variability from person to person and could account for some of those rare 'out-of-sight' mutations with moderate penetrance that GWAS can't pick up. Many CNVs go undetected because they don't alter SNP sequences. Duplicated regions can also be difficult to sequence.

A standard technology for uncovering CNVs is array comparative genomic hybridization, in which scientists examine how genetic material from different individuals hybridizes to a microarray. If certain spots on an array pick up more or less DNA, it could indicate that there's a CNV. This and several other techniques are being tested by a consortium called the Copy Number Variation Project, run out of the Wellcome Trust Sanger Institute in Cambridge, UK. The consortium is dedicated to characterizing as many CNVs as possible so that associations can be made between them and diseases. McCarthy says that the role hidden CNVs have in heritability "should play out in the next six months to a year". But Goldstein argues that current technologies will miss many of the smaller CNVs, from 50 base pairs down to repeats of just two bases. "All we'll have verification of is the big whopping CNVs that are identifiable, and they clearly do not account for much of the missing heritability."

In underground networks
Most genes work together with close partners, and it is possible that the effects of one on heritability cannot be found without knowing the effects of the others. This is an example of epistasis, in which one gene masks the effect of another, or where several genes work together. Two genes may each add a centimetre to height on their own, for example, but together they could add five. GWAS don't cope with epistasis very well, and efforts to find these interactions usually require good up-front guesses about the interacting partners.

Joseph Nadeau, a geneticist at Case Western Reserve University in Cleveland, Ohio, says that 'modifier' genes act even in some straightforward single-gene diseases. "That's a simple kind of epistasis," he says. Cystic fibrosis, for example, is usually caused by mutations in one gene, *CFTR*, yet can vary greatly in symptoms and severity. The suspicion has been that modifier genes are one cause of this variability.

But despite the years of study, researchers still struggle to pin down these genes. "People haven't modelled truly the effect of epistasis," says population geneticist Sarah Tishkoff at the University of Pennsylvania in Philadelphia.

It's no surprise that genetics is more complicated than one gene, one phenotype, or even several genes, one phenotype, but it's humbling to realize how much more complex
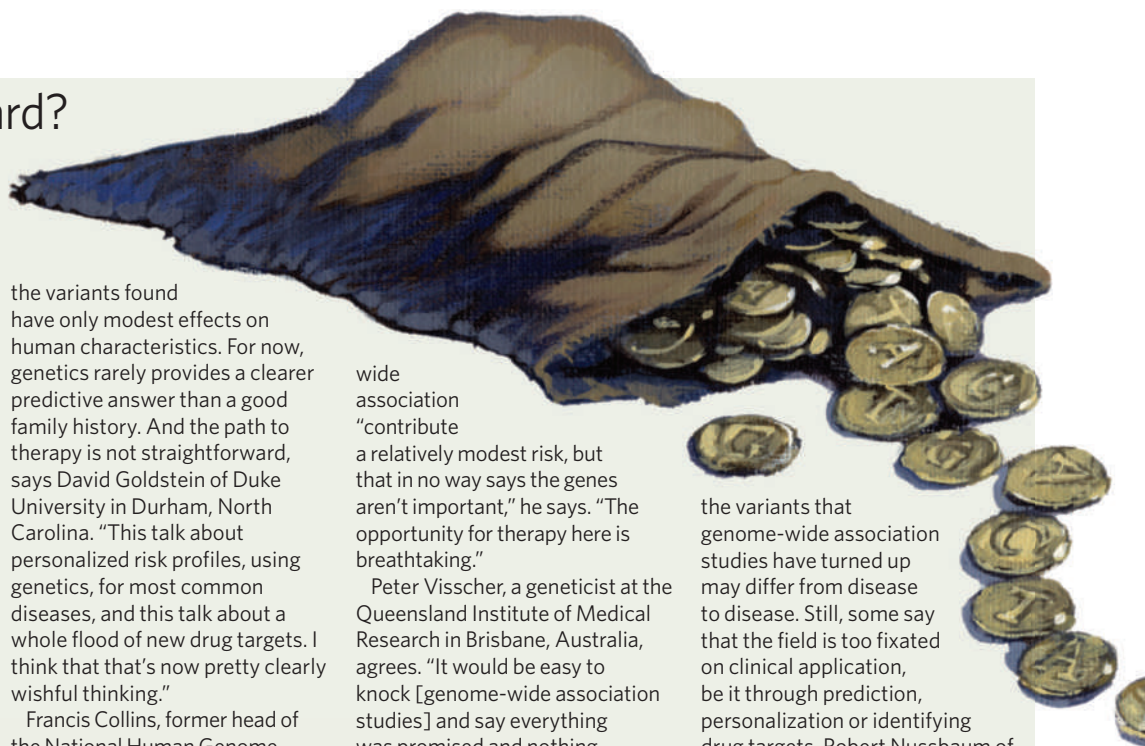
things are starting to look. In a now classic study[10], Kruglyak and his colleagues found that expression of most yeast genes is controlled by several variants, often more than five. To fill in all the heritability blanks, researchers may need better and more varied models of the entire network of genes and regulatory sequences, and of how they act together to produce a phenotype. At some point this process starts to look more like systems biology, and researchers are already applying systems methods to humans and other organisms (see page 26). "What we're learning from these studies is that we need to think about the more complex of the complex models rather than the more simple of the complex models," Kruglyak says.

The great beyond

What if heritability estimates were wrong in the first place? Heritability of height was initially measured by taking the mean height of parents and comparing that value to the adult height of their offspring. As the average heights of parents increase, researchers found, so too does the average height of their children, hence the calculated 80–90% heritability.

Environment, especially factors such as nutrients or toxins present during important growth phases, can affect the mean height of a population considerably — but researchers have controlled for environment in estimates of heritability by, for example, comparing genetically identical twins raised together with those raised apart. Most researchers are confident that the heritability estimates are sound. "I don't think anyone's going to say that the heritability of height is 10% and let environment get you closer to the answer," Kruglyak says. "I don't think you can explain it away."

But there are lingering doubts about how precisely environment has been accounted for in heritability studies. Adverse experiences *in utero* could lead to lifelong health disparities, according to David Barker from the University of Southampton, UK, and yet a shared womb is an aspect of the environment that would not be factored into such studies. "Heritability estimates are basi-

cally what clusters in families, and environment clusters in families," says Manolio.

Epigenetics, changes in gene expression that are inherited but not caused by changes in genetic sequence, confuses things further. Feeding a mouse a certain diet, for example, can alter the coat colour not only in its children, but also in its children's children[11]. Here, the expression of a coat-colour gene is controlled by a type of DNA modification called methylation, but it's not completely clear how that methylation pattern is 'remembered' by the next generation. The idea that grandma's environment could affect future generations is controversial — and such effects would have been included in the heritability normally attributed to genes.

"This complicates everything," says Nadeau. "How do we sort out what great-grandfather and great-grandmother were exposed to when they were young and having children?" Model organisms might help. Nadeau has investigated testicular germ-cell tumours in mice that are analogous to a highly heritable cancer in humans. His group found that the effects of one weak, cancer-promoting gene, $Dnd1^{Ter}$, are greatly enhanced by several other gene variants, and the boosted effects are passed on even if the genes that cause them are not[12].

"It's presumably transmitting its presence in some epigenetic way," says Nadeau. The mechanisms by which epigenetic inheritance might work are still disputed, though; marks such as methylation that direct gene expression during someone's life seem to be wiped clean in a new embryo. One possible explanation for Nadeau's observation, he says, is that RNA is being inherited alongside DNA through sperm or eggs.

Collins is not convinced that epigenetics will play a big part in missing heritability in humans. "It just doesn't look likely outside of one or two examples to suggest that this is the case." Nadeau disagrees. "It's hard to imagine that every other organism works one way and humans are the exception," he says.

Lost in diagnosis

There is a nagging worry as researchers hunt for heritability: that common diseases might not, in fact, be common. Medicine tries hard to lump together a complex collection of symptoms and call it a disease. But if thousands of rare genetic variants contribute to a single disease, and the genetic underpinnings can vary radically for different people, how common is it? Are these, in fact, different diseases?

GWAS could actually be proving so difficult because researchers are seeking shared susceptibility genes in a group of people who may share few, if any. And yet without a more refined understanding of genetics, it could be impossible to categorize them any better. "It may be rare variants, common disease. And that's kind of scary to people because it's much, much harder to find those," says Tishkoff.

There could be scarier and more intractable reasons for unaccounted-for heritability that are not even being discussed. "It's a possibility that there's something we just don't fundamentally understand," Kruglyak says. "That it's so different from what we're thinking about that we're not thinking about it yet."

Still the mystery continues to draw its sleuths, for Kruglyak as for many other basic-research scientists. "You have this clear, tangible phenomenon in which children resemble their parents," he says. "Despite what students get told in elementary-school science, we just don't know how that works." ■

Brendan Maher is a Features editor for *Nature*.

1. Visscher, P. M. *Nature Genet.* **40,** 489–490 (2008).
2. Weedon, M. N. *et al. Nature Genet.* **40,** 575–583 (2008).
3. Lettre, G. *et al. Nature Genet.* **40,** 584–591 (2008).
4. Gudbjartsson, D. F. *et al. Nature Genet.* **40,** 609–615 (2008).
5. Sullivan, P. F. *PLoS Med.* **2,** e212 (2005).
6. Freitag, C. M. *Mol. Psychiatr.* **12,** 2–22 (2007).
7. Romeo, S. *et al. Nature Genet.* **39,** 513–516 (2007).
8. Stefansson, H. *et al. Nature* **455,** 232–237 (2008).
9. The International Schizophrenia Consortium *Nature* **455,** 237–241 (2008).
10. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. *Science* **296,** 752–755 (2002).
11. Waterland, R. A. & Jirtle, R. L. *Mol. Cell. Biol.* **23,** 5293–5300 (2003).
12. Lam, M. Y., Heaney, J. D., Youngren, K. K., Kawasoe, J. H. & Nadeau, J. H. *Hum. Mol. Genet.* **16,** 2233–2240 (2007).

21
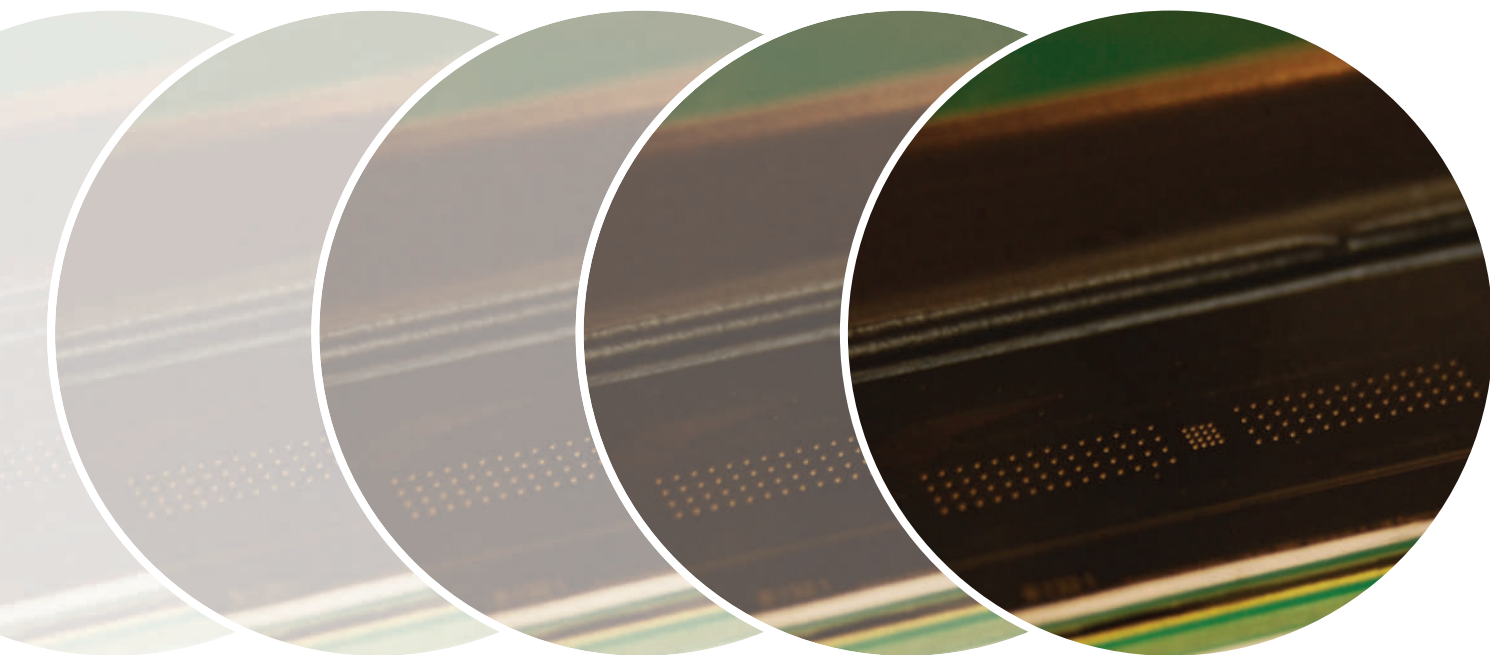
STANDARD AND PORES

Could the next generation of genetic sequencing machines be built from a collection of miniscule holes? **Katharine Sanderson** reports.

DNA sequencing is a technology on the move. In April, 454 Life Science, based in Branford, Connecticut, sequenced the entire genome of James Watson in two months for less than US$1 million[1]. In this issue, Illumina, based in San Diego, California, reports the sequence of a human genome obtained for a quarter of that price and in eight weeks[2]. Companies are positioning themselves aggressively to go further, faster and cheaper.

Many consider the ideal technology to be 'single molecule' sequencing, which reads from individual DNA fragments without the need for amplification and the risk of introducing errors. Pacific Biosciences, based in Menlo Park, California, has placed itself centre stage, promising to deliver such a service by watching enzymes build DNA base by fluorescently tagged base. But the single-molecule technology that the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, has invested most in is nanopore sequencing, in which DNA is read as it threads through a tiny hole. The technique has received $40 million of a total of $68 million spent in the institute's drive to generate human genomes for $1,000. $4.2 million of that went to Hagan Bayley, a chemical biologist at the University of Oxford, UK, to back research that forms the basis of Oxford Nanopore Technologies, the company he founded, and the one that is closest to making a working nanopore sequencer.

Jeffrey Schloss, NHGRI programme director of technology development, says that nanopore sequencing is the only method the institute has supported so far that has the potential to sequence DNA directly from cells without amplification, modification or use of expensive reagents such as fluorescent tags. Oxford Nanopore Technologies's chief executive Gordon Sanghera says that he would like his technology to "dominate the world, ultimately". But Sanghera faces stiff competition. Pacific Biosciences, and Complete Genomics in Mountain View, California, are just two of the companies that have announced their ambition to become the chief provider of genetic sequencing. There is still scepticism in the scientific community about whether nanopore sequencing can deliver, says Schloss, and there is a simple reason: "Pacific Biosciences and Complete Genomics have both sequenced some DNA. Nanopores have not."

One of the first suggestions that nanopores could form the basis for DNA sequencing came in 1996, when a team led by Daniel Branton, a biophysicist at Harvard University, showed that the presence of DNA could be detected as it passed through a pore by the interruption in the flow of ions through the aperture[3].

The pores, made from a ring of seven α-haemolysin membrane proteins, are the same as those that the infectious bacterium *Staphylococcus aureus* pushes into the membranes of other cells in order to create damaging holes. Branton's result suggested that the identity of each of the four bases traversing the hole might be revealed by distinctive changes in ion flow, which can be read as an electrical signal.

From small beginnings

Bayley and Sanghera founded the company in 2005 to develop nanopores as sensor systems for DNA and other molecules, but the company quickly decided to focus on DNA sequencing. Bayley provided 20 years of experience studying nanopores and Sanghera, who had previously worked for Abbott Laboratories, the business know-how. Of the $35 million that has been raised to finance the company, all from private and institutional investors, $20 million came in a financing round in March this year.

In 2006, Bayley showed that the distinction between bases could be made when each was held in place in the nanopore for long enough[4].
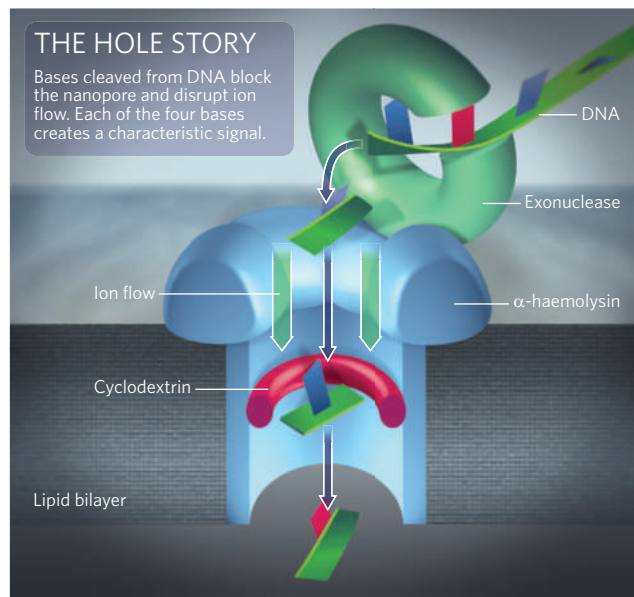
Oxford Nanopore Technologies's 128-pore chips (top) build on work by Hagan Bayley (above).

"The breakthrough is that one free nucleotide gives a distinguishable signal," says Tim Harris, from the applied physics and instrumentation group at the Howard Hughes Medical Institute's Janelia Farm Research Campus in Ashburn, Virginia.

DNA cannot, for now, be run continuously through the nanopore, partly because of the need to hold each base in the pore long enough to disrupt the flow of ions. So, to do their sequence detection, Bayley's group has used genetic engineering and chemistry to make two alterations. At the pore's mouth, the team placed an exonuclease, an enzyme that grabs the ends of a DNA molecule from a solution running over the top. The enzyme then severs each base and directs it into the hole (see graphic). At the other end of the pore, the group inserted a cyclodextrin plug, a ring-shaped molecule that narrows the neck. The passing bases have to squeeze through this plug and, as they do so, a phosphate group on the nucleotide briefly binds the cyclodextrin and blocks the pore. Because the bases are different sizes, they sit within the cyclodextrin for different lengths of time, and fill it to different extents, giving characteristic readouts for each base.

"The advantage of this technique is, first of all, it's a single-molecule technique, so you don't have to amplify or clone your DNA," says Bayley. There are no fluorescent tags and, in theory, minimal sample preparation. "Also you're directly sequencing the genomic DNA, so, in principle, as well as just getting the four bases you should be able to get modified bases," Bayley adds. Oxford Nanopore Technologies says it has unpublished data showing that the system can better discriminate between the four bases and detect 5-methylcytosine, a chemically altered version of cytosine that is commonly involved in gene regulation.

In May this year, the company decided that its technology had advanced far enough to announce its intention to develop a next-generation sequencing system. The company had also been quietly vacuuming up the intellectual-property rights from some of the leading nanopore research teams, signing licensing deals with leaders in the field such as Branton, and David Deamer and Mark Akeson at the University of California, Santa Cruz. "They're eliminating their competition," says Harold Swerdlow, head of sequencing technology at the Wellcome Trust Sanger Institute in Cambridge, UK.

Key questions

The part of the project that the company is reluctant to talk about is the bit that everyone most wants to know: how this will be scaled up into a working, multichannel sequencer. How many working pores could be used in parallel, and how quickly would it sequence a DNA strand? And crucially, when will sequencing data be made available?

Early prototypes in the company's lab look far from complete. A ten-square-centimetre chip, capable of holding 128 pores that will sequence different DNA fragments, sprouts plastic tubing that delivers the samples and naked wiring that connects to an electronics box. But those at the company are tight-lipped about the details of the final product, how it might work and when. They say that they do not want to oversell themselves by making a

THE HOLE STORY
Bases cleaved from DNA block the nanopore and disrupt ion flow. Each of the four bases creates a characteristic signal.

DNA

Exonuclease

Ion flow

α-haemolysin

Cyclodextrin

Lipid bilayer

ACGT spells hype

When Complete Genomics, based in Mountain View, California, announced in early October that it would "offer complete human genomes for $5,000", Jim Hudson wondered if it could live up to its claim. So when Hudson, the president and co-founder of the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama, met company representatives at a meeting later that week, he offered them an envelope stuffed with several thousand dollars cash in exchange for his sequence. "I said, 'I want one'," Hudson says.

The representatives declined the envelope, explaining that the company isn't actually taking orders at that price — which was promised by 2009. Hudson and many other scientists are still left wondering whether the company can live up to its promises.

A gap has opened up between the claims made by companies offering the next generation of cheap, ultrafast DNA sequencing and the data to back them up. The $1,000 genome is expected to attract biotechnology and pharmaceutical firms to a sequencing market that has so far been limited to academic centres. Sequencing companies wanting a share of this new market are walking a delicate balance between gaining investor interest and avoiding damaging their credibility with unsupportable claims.

In December 2005, Helicos Biosciences of Cambridge, Massachusetts, announced that it had sequenced the genome of a bacteriophage using a technology that could potentially sequence entire human genomes in one day. The company then raised $40 million in venture capital and nearly $46 million in its May 2007 initial public offering. But problems with its machines then caused long delays while competitors announced that they had sequenced human genomes. Helicos still seems far from achieving this and has sold few machines, causing its share price to drop from a high of more than $17 this January to a low of 60 cents last month. Chief scientific officer Patrice Milos says the company has ironed out its problems and will present new data at the Advances in Genome Biology and Technology (AGBT) meeting in February 2009.

Helicos has also been hurt by the debut of another big name in this field. Pacific Biosciences of Menlo Park, California, has been developing a new sequencing technology based on DNA polymerase, an enzyme that copies DNA, and made a splash by presenting data at the AGBT meeting in Marco Island, Florida, this February. There, the company showed that it had sequenced pieces of DNA about 150-base-pairs long — a tiny step towards its eventual goal of sequencing entire human genomes in less than 10 minutes, but a tangible piece of data that has built up the company's credibility among

specific prediction that they will do X in Y time, and then disappointing or surpassing those expectations. "There's a danger for a company like this to come out too soon," Sanghera says. "It's a very difficult commercial strategy." (see 'ACGT spells hype').

Swerdlow is talking with all of the new companies. "It's quite difficult to decide who's telling the truth," he says, "It's all hearsay to some extent." He remains optimistic but unconvinced about Oxford Nanopore Technologies. His concern is whether the reagents needed to run a sequence might break down the biological pore in some way.

> **"There's danger for a company like this to come out too soon."**
> — Gordon Sanghera

"I do think that there is some scepticism about direct nanopore sequencing," says Barrett Bready, chief executive of sequencing start-up NABsys in Providence, Rhode Island. He says this scepticism is based on the inherent difficulty of the problem. "The four bases actually differ by only a few atoms. These differences must be detected in the face of noise from various sources."

NABsys, formed in 2004 by Xinsheng Sean Ling, a physicist at Brown University in Providence, is also pursuing nanopore sequencing, but seems to be further from a working machine than its Oxford rival. In 2007, Ling and John Oliver, another NABsys scientist, received two NHGRI sequencing grants worth $1.32 million in total. The method is based on a silicon chip dotted with synthetic nanopores. Through these pores pass 100,000-base-long fragments of genomic DNA that have six-base-long probes attached to them at intervals. The method uses a library of probes, each having a different, but known, sequence. As the DNA passes through the pore, the points at which a probe is attached can be detected from the current in the chip. The time gaps between those current readings allows the location of the probes to be determined. Once lots of fragments are probed in this way, a picture of the entire genome can be put together from these sequences. But Bayley is dubious. "You can engineer [proteins] with angstrom precision, which you simply can't do with a pore in plastic or silicon nitride at this point." And Harris says that NABsys's sample preparation, which involves reengineering the DNA, is clunky. "This seems like an improbably gymnastic sample process for something that has to be fast, and essentially free."

George Church, a molecular geneticist at Harvard University, whose work has also been licensed by Oxford Nanopore Technologies, thinks that the sequencing race will be won by whichever company has the lowest instrument cost and the highest throughput per instrument. Sequencing methods that rely on a digital camera to record colour changes from fluorestly tagged bases — such as Pacific Biosciences' technology — are winning that race over nanopores, he says. "Digital cameras are capable of collecting millions of bits of information at close to the maximum data-flow rate that a PC can handle." The cost of these cameras has dropped because of huge consumer use, says Church. "It does not seem to be a similar case for massively parallel ion-channel monitors."

Schloss says that the NHGRI views nanopore sequencing as a long-term goal. "We expected, when we launched the programme in 2004, that it might well take ten years to achieve the goal of using nanopores for sequencing DNA." Sanghera has no such reservations. "Our products are going to be so good that we're just going to let the technical data speak for itself. All things will flow from that." ∎

Katharine Sanderson is a reporter for *Nature* based in London.

1. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
2. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
3. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. *Proc. Natl Acad. Sci. USA* **93,** 13770–13773 (1996).
4. Astier, Y., Braha, O. & Bayley, H. *J. Am. Chem. Soc.* **128,** 1705–1710 (2006).

See Editorial, page 1.

J. DEPCZYK

---

scientists — and helped it to raise $100 million in venture capital as of July 2008.

Complete Genomics' surprise entrance was more audacious than anything the field has seen so far. The company pledged to cut sequencing costs to $5,000 in six to nine months, but hasn't offered any data, annoying some potential customers. Richard Gibbs, director of the Baylor College of Medicine genome sequencing centre in Houston, Texas, says he had just convinced a private donor to fund a large cancer genome sequencing study when Complete Genomics' announcement hit the press, prompting the donor to ask why Gibbs needed $350,000 per genome when a new company was sequencing genomes at a fraction of that cost. "I'm sure you all appreciate the dilemma that poses," Gibbs told scientists at an October meeting at Cold Spring Harbor Laboratory, New York.

Clifford Reid, the chief executive of Complete Genomics, says the company will also present data at February's AGBT conference, showing sequences of parents and children from a study with Lee Hood at the Institute for Systems Biology in Seattle, Washington.

Other sources told *Nature* that Complete Genomics is currently charging as much as $20,000 per genome, with a minimum order of five genomes and a six-month wait time. Reid says that current prices are based on the size of each order and will come down as the company's capacity grows.

"I don't think we've created unrealistic expectations," Reid says. "But I think we've put a lot of people in a difficult position [because] this is a disruptive technology, and one of the great challenges in the scientific and medical community is coping with this kind of disruption." Scientists, for their part, say they have little disruption to cope with until they see the sequence.
Erika Check Hayden
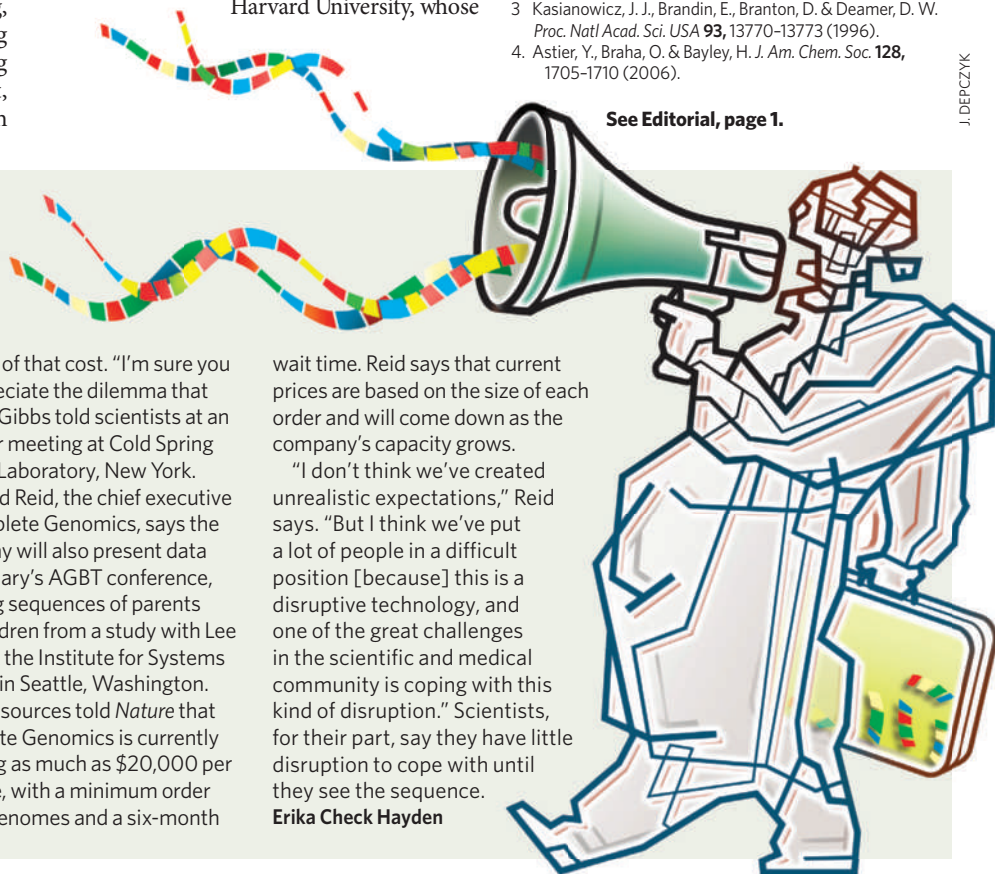
A DISRUPTIVE PERSONALITY, DISRUPTED

Eric Schadt revels in making people uncomfortable with his science. **Bryn Nelson** reports how the bioinformatics rabble-rouser hopes to charge ahead in the face of his company's disintegration.

In need of an escape from the mental gymnastics of hardcore genome analysis, Eric Schadt, executive scientific director of research genetics for Rosetta Inpharmatics, is clear about what works for him — careering down a steep mountainside on a snowboard. "You can't sort of ease your way down a hill," he says over breakfast near the company's headquarters in Seattle, Washington. "It's either all, or nothing."

That fearless approach may be tested after a bombshell announcement last month. Rosetta, which has been on a head-turning run for most of the past decade, now finds itself in mid-air, hoping to make a landing that could be very tricky indeed. Merck & Co., which bought the biotech company in 2001 and operated it as a subsidiary, announced on 22 October that it will close down most of the Seattle operations by December 2009, transferring Schadt and a number of his team to its Boston research centre.

The move comes as part of a global reorganization in the face of slumping drug sales, and includes cutting more than 7,200 jobs worldwide. It could be worse. Reni Benjamin, a senior biotechnology analyst and managing director for investment bank Rodman & Renshaw in New York, says that given the current economic climate for pharmaceuticals, sub-sidiaries such as Rosetta have no guarantee of survival. "They could just cut the entire thing and call it a day if they didn't think it was important."

What's likely to save Rosetta from complete oblivion is Schadt's trend-setting science of integrated genomics, which uncovers disease mechanisms by revealing vast networks of gene interactions. Genome-wide association studies, which have been a favoured technology for finding disease-linked genes over the past several years, seek out associations between a disease state and a genetic variant. Although the studies have turned up hundreds of variants associated with disease, they can detect only independent effects of individual genes, which means they might miss a lot (see page 18).

Like genome-wide association studies, Schadt's approach tries to correlate variations in DNA with some observable complex trait, such as a disease, in a population. But Schadt and his group add a third factor: gene-expression levels, as measured by microarrays. They then use the data to build models of how the three factors — variations in the genotype, variations in the disease, and variations in gene expression — are related.

Some relations are straightforward: a gene variant has a direct effect on expression, and that has a direct effect on the disease trait. But it is also possible for a particular genetic variation to be linked to a disease trait that, in turn, alters the expression of some other gene. And then there are cases in which a genetic variation is linked to both a disease and an unrelated change in gene expression. The complexity of all this goes through the roof when genes start interacting: the models explode into networks of interconnected elements. But these network models allows Schadt and his colleagues to identify and validate associations between genes and disease — and to predict how perturbing the system, with a drug or genetic mutation, will affect expression and disease.

> "Their whole lives were 'What the heck is that gene?'; my point was 'Most of those genes aren't even druggable'."
> — Eric Schadt

Success story?
So far, company officials say, the strategy has worked well. Of the 52 compounds in Merck's clinical-trial pipeline in 2006, ten entered through Rosetta's efforts. Now, according to Stephen Friend, a senior vice-president at Merck and Rosetta's co-founder, the approach is so integrated, it's hard to tell what originated from Schadt's team. Something that important

is not going to be axed just to control costs, says Benjamin. "If 20 to 25% of a company's pipeline is being generated from one particular platform, you would have more significant clues than 'streamlining operations' if they didn't like what they were seeing."

On Rosetta's fourth floor, there's all sorts of stuff that may not survive the move east. A cheeky monument built from lab equipment including discarded flasks, racks and pipettes, entitled 'Don't Mess With Us, We're Scientists', sits near display cases showing off the company's early innovations such as its microarray spotter prototypes. One recent acquisition is likely to stay on the packing list. The Hamilton MicroLab Star, a custom-designed platform hosting three interconnected robots, represents the next generation for gene-expression profiling, offering a nearly seamless start-to-finish automated process.

Upstairs, John Dey, the company's UNIX operations manager, lifts a floor panel to reveal a sea of blue cables and a rough visual gauge of the computing power housed within the company's nerve centre. The cluster began with eight computer nodes in 1998. Now it boasts 1,000 connected by more than 16 kilometres of cable. The computational muscle in Boston will have to be bigger still. Dey shakes his head when asked whether he's confident about keeping up with storage demands. "Oh, it's going to explode," he says.

Schadt's team needs this state-of-the-art equipment for what Schadt calls one of the biggest looming bottlenecks for biology in the next ten years: understanding the effects of many genes interacting at once. For genome-wide association studies, the question is fairly straightforward from a computational perspective. But testing biology's dizzying network of interactions in a holistic way, Schadt says, requires the computational power generally reserved for climate modelling and astrophysics.

This aggressive approach rightfully has made specialists uneasy. Friend says integrated genomics met with an initial 'show me'-style distrust. With any new technology, he says, an evangelical wave first oversells it and turns almost everyone into a sceptic.

Schadt clearly revels in proving the critics wrong. In May, Rosetta led a study of gene expression in the human liver and found more than 6,000 associations between single nucleotide polymorphism genotypes and gene expression traits[1]. Although many of the expressed genes were already implicated in

human disease, other connections were newly exposed as suspect. The study's integrated data, for example, pointed to *RPS26* instead of the previously favoured *ERBB3* as the most likely susceptibility gene in a genomic region associated with type 1 diabetes.

The layering approach by which Rosetta constructs a complex network has garnered its own share of detractors. "To some of these naysayers, you have a big magic black box where you pour in everything and out comes a drug target, and that doesn't sound like science," says Trey Ideker, an associate professor of bioengineering at the University of California, San Diego, who is collaborating with Schadt on a review detailing the use of systems and network modelling for drug development and health care. "But if you look under the hood, what Eric is doing is absolutely sound — it's the sheer complexity that is overwhelming."

Friends reunited

Schadt often winds his critics up, of course. "The network stuff still makes a lot of people uneasy," he says. "People don't want things to be that complicated." He smiles as he recalls his admittedly "inflammatory" talks a few years ago, in which he basically told other scientists: forget genes and focus on what genetic perturbations are doing to the whole system. "Human geneticists just hated it when I would say stuff like that," he says. "You know, their whole lives were, 'What the heck is that gene?' and my whole point was,

> "What Eric is doing is absolutely sound; it's the sheer complexity that is overwhelming."
> — Trey Ideker

'Well, first of all, most of those genes aren't even druggable'."

Most of Rosetta's methods have all but ignored the question of gene identity initially and instead tracked disease-associated hiccups in a genetic network. The approach led to the 2005 publication in *Nature Genetics* of a study laying out the general integrative strategy — something Schadt counts among his proudest accomplishments[2]. "It wasn't just the intellectual satisfaction, it was that everybody — nearly everybody — was saying, 'Nah, it's not going to work'," he recalls. He looks back almost wistfully on those earlier fights. "Because our work has got so successful now, I don't feel that people push back enough," he says. "It's almost too accepting."

The troubled economic landscape could provide plenty of pushback to Schadt's resource-intensive approach, but he seems unfazed by unanticipated changes. Unplanned course corrections have defined his past. Raised in rural Michigan by a conservative family that placed little value on education, Schadt joined the US Air Force after high school and enrolled in its elite pararescue programme, sometimes called 'superman school' because of its gruelling training regimen. During one exercise, Schadt dislocated his shoulder so badly he required surgery, ending his rescue career. His superiors, though, noticed his high test scores and asked about college. It had never occurred to him.

With financial assistance from the military, he gravitated towards mathematics and computer science at California Polytechnic State University in San Luis Obispo. "It just opened up a brand new world." It wasn't until he earned
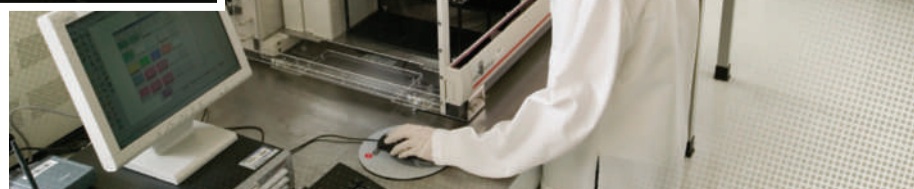
A Merck scientist prepares samples for hybridization to DNA microarrays.

MERCK & CO.

his master's degree in pure mathematics from the University of California, Davis, that biology first caught his eye. Schadt began attending biology seminars and met Kenneth Lange, now chairman of the human-genetics department at the University of California, Los Angeles. Lange encouraged Schadt to pursue a curriculum in biomathematics and eventually became his PhD adviser, guiding him to a degree that integrated molecular biology and applied mathematics. "It just all made sense," Schadt says.

By the time Merck began courting Rosetta in 2001, Schadt had become a key member of the team behind a seminal annotation of the human genome[3]. Annotation aided by gene-expression microarrays was fast gaining attention, Schadt says, "and we just thought we were going to make it big".

The company, in fact, was on the verge of agreeing to a lucrative contract to expand on its annotation work. "The day we were supposed to sign that deal, Stephen Friend comes in and says, 'We're not going to sign,'" Schadt recalls. "So I said, 'Do you have a better $20-million deal?'" Merck ended up buying Rosetta in a stock swap valued at about $620 million.

Next year, Schadt and Friend will be reunited in Boston, where Merck is hoping for a better synergy by co-locating its molecular-profiling and integrative-genomics research with Friend's oncology group, as well as the researchers working on bone and respiratory diseases, arthritis and analgesia. "By providing a more comprehensive view of the numbers of genes that may be causally related to disease," Schadt says, "we can actually use the networks to prioritize those targets."

Identifying and halting a drug programme headed for trouble can be just as important, and Schadt cites the *Ppm1l* gene as a perfect example. In March, his team found an obesity connection for *Ppm1l* — which encodes a newly discovered but poorly characterized protein phosphatase — and two other previously unlinked genes. But the group found that *Ppm1l* also modulated traits for diabetes and high blood pressure[4]. "What we showed is when you knock the gene down, you improve your diabetes profile or you become less insulin resistant, but you also gain weight." Even worse, knocking down the gene also increased blood pressure. Making someone heavier and with higher blood pressure in exchange for a lower diabetes risk, the company concluded, wasn't worth the trade-off

— something it might have missed by taking a narrower focus.

Thomas Gingeras, head of functional genomics at Cold Spring Harbor Laboratory in New York, says Schadt should be commended for embracing a systematic approach to teasing out the molecular networks in cells. But he worries about the initial downside to such an ambitious plan. "The concern I have is whether the information we have to be able to put together this network is patchy and, in some cases, unreliable; and, even if it is accurate, whether we have the right interpretation of those data," he says. By focusing on the low-hanging fruit within the network, might researchers be losing sight of more important non-coding elements that aren't yet within reach?

A legacy lives on

Schadt is confident that his team's models can adapt as a wider range of information comes online, including forthcoming studies that will incorporate data from genome-wide screens of small-molecule metabolites and protein–protein interactions. He's particularly enthusiastic about several big pilot projects that are resequencing entire transcriptomes for hundreds of mice and humans. They offer a way to ask whether largely unknown non-coding RNAs may be central players in the protein-coding network.

Ritsert Jansen, a bioinformatics expert at the University of Groningen in the Netherlands, says Schadt's ability to use high-throughput screens and work with large patient

> **"It wasn't just the intellectual satisfaction, it was that everybody was saying, 'Nah, it's not going to work'."**
> — Eric Schadt

populations is a big step closer to explaining the causality of complex traits. Jansen, who works on molecular networks in the *Arabidopsis* plant model, says Schadt's work has so far suggested that a few very influential DNA variations are critical for linking genotype with phenotype. The expanding repertoire of 'omics-based studies, including epigenomics, should lead to more dynamic models that zero in on the most important perturbations, he says.

One of the next phases in that progression will add in more clinical information from Merck collaborators such as the Moffitt Cancer Center and Research Institute in Tampa, Florida. Every cancer patient giving informed consent will have multiple tissue samples collected and analysed, with an eye towards charting tissue-to-tissue communication networks.

And then what? Rosetta's legacy may be in providing Merck with a mastery of biological information and superior predictive models of disease risk, drug development and patient response. But ultimately, Schadt says, consumer genome-testing products, such as those provided by 23andMe in Mountain View, California, Navigenics in Redwood Shores, California, and deCODEme of Reykjavik, Iceland, will take the lead in solving perhaps the problem of the century: translating all that information for the people who need to know what it means, be they doctors or parents. "The next ten years are going to be an amazing ride to see how this all plays out," he says.

Despite Schadt's sentiment that "change is good from time to time", he concedes that keeping his team together and focused on its research during the move east will be challenging. But his overall mission, he says, remains unchanged. In an e-mail to *Nature* shortly after Merck announced it would close Rosetta and two other research units in Japan and Italy, Schadt wrote: "It is very gratifying that our work is so highly valued within Merck and that as a result it will become even more integrated within Merck's research." The bottom line: "everything is continuing, we will just be doing it in Boston instead of Seattle". ∎
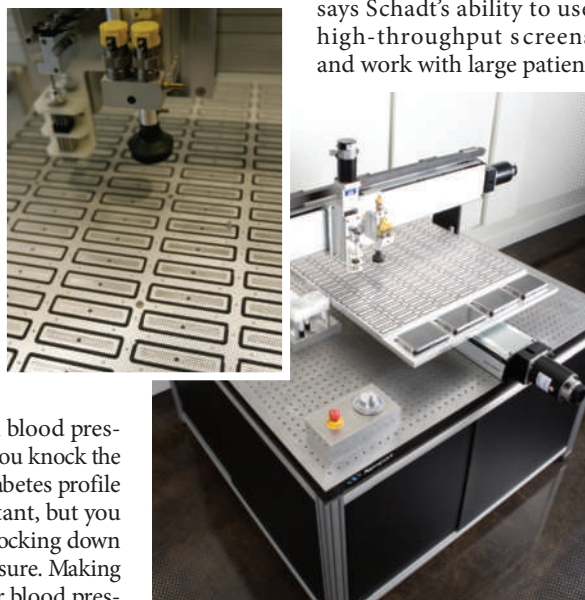
Bryn Nelson is a freelance science writer based in Seattle, Washington.

1. Schadt, E. E. *et al. PLoS Biol.* **6,** e107 (2008).
2. Schadt, E. E. *et al. Nature Genet.* **37,** 710–717 (2005).
3. Shoemaker, D. D. *et al. Nature* **409,** 922–927 (2001).
4. Chen, Y. *et al. Nature* **452,** 429–435 (2008).

MERCK & CO.

MERCK & CO.

Rosetta's 1997 technology for studying yeast expression was quickly replaced by newer tools.

CORRESPONDENCE

Research rewards are worth the effort for multitasking mothers

SIR — The reasons women drop out of science are complex, and Timothy Roper and Larissa Conradt have hit on an important factor in their Correspondence 'Childcare not enough to make a science career family-friendly' (*Nature* **455,** 1029; 2008). However, I don't see encouraging more women into science as either pointless or unethical.

Careers in science can offer enormous rewards to women. Moving into an academic environment has provided great opportunities for me as a mother, owing to its flexibility. I am now measured largely on my productivity, and my ability to multitask — honed by motherhood — is an asset as I juggle research, administrative duties and teaching.

I have worked in the male-dominated field of Antarctic research for the past 15 years, and I run a research programme looking at climatic warming impacts on the top predators, leopard seals. This work has been successful, thanks to my scientific team — which, incidentally, is mainly composed of women. As the mother of two children under the age of six, I suspect that a large part of my success has been due to the enduring support of my partner. I'm not going to pretend that it has been plain sailing, but I wouldn't have done it any differently.

Let's stop asking why there are so few women in science. Instead, let's turn the question round to ask how those who made it actually got there.

As scientists, we are skilled strategists, overseeing the conception of a new research initiative, then the project's gestation and its birth as a peer-reviewed article. These planning skills also sustain our lives outside the lab.

To those women embarking on the journey, I would say that it is not a road for everyone — but if, like me, you have a burning passion for your research, I would encourage you whole-heartedly to pursue it. It's a long journey, so pace yourself and plan — including your home life and time with your family in your plan. Sometimes you need to step back a little in order to move forwards.

Tracey L. Rogers Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney 2052, Australia
e-mail: tracey.rogers@unsw.edu.au

Readers are welcome to comment at http://tinyurl.com/56mavj

What is nature, if it's more than just a place without people?

SIR — Your Editorial 'Handle with care' (*Nature* **455,** 263–264; 2008) notes that many people define 'nature' as a place without people, and that this would suggest that nature is best protected by keeping humans far away. You question the value of this negative definition, arguing that "if nature is defined as a landscape uninfluenced by humankind, then there is no nature on the planet at all".

This may be true. However, if we define nature as including humankind, the concept becomes so all-encompassing as to be practically useless.

As an ecologist, I consider humans to be embedded in nature rather than separate from it. This relationship does not disappear in an urban environment. For example, the food you eat, the paper you read and the energy you consume are all products of multiple interacting organisms and ecosystem services. But if we adopt this inclusive definition, it becomes impossible to identify anything on the planet that is not a part of 'nature'. In this case, an atom bomb becomes as 'natural' as an anthill.

A dilemma therefore arises. If nature is somewhere that humans are not, we lose sight of the fact that we are just another species intimately intertwined in the complex web of biological systems on this planet. However, if we place ourselves within a definition of nature, the definition then becomes essentially meaningless by extending to everything on Earth.

Your Editorial comments that "Nature doesn't have to end if we stop defining it by humankind's absence". The problem is that once we no longer define nature by our absence, the concept has no end.

Is there a better definition of nature?

Fern Wickson Centre for the Study of the Sciences and the Humanities, University of Bergen, PO Box 7805, 5020 Bergen, Norway
e-mail: fern.wickson@svt.uib.no

Progress being made on standards for use in data sharing

SIR — In his Correspondence on data sharing ('Big data: open-source format needed to aid wiki collaboration' *Nature* **455**, 461; 2008), Tin-Lap Lee points out "there is currently no de facto standard on pathway-data format, which severely limits data portability". Although this statement is correct, there are three particular standards in use — BioPAX (www.biopax.org), CellML (www.cellml.org) and the Systems Biology Markup Language (www.sbml.org) — that all serve this purpose.

These standards can provide annotations based on appropriate ontologies. In other words, they provide an indexing system that gives unambiguous representations of the entities that they describe, thereby avoiding the problem of synonyms (see M. J. Herrgård *et al. Nature Biotechnol.* **26**, 1155–1160; 2008).

MIRIAM is a community recommendation for minimal information to be reported about models and can be used with any of the three standards (N. Le Novère *et al. Nature Biotechnol.* **23**, 1509–1515; 2005).

I would strongly recommend that everyone with an interest in sharing models should do so using one or more of these formats.

Douglas Bruce Kell Manchester Centre for Integrative Systems Biology, School of Chemistry, and Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
e-mail: dbk@manchester.ac.uk

One-year practical course proves a launch pad for PhDs

SIR — In response to Cristina Banks-Leite's comments in Correspondence ('More ground work needed to prepare students for PhDs' *Nature* **455,** 285; 2008) — I couldn't agree more.

I secured a PhD scholarship straight from being an undergraduate. In retrospect, I believe that both I and my PhD supervisor would have benefited enormously if I had gone through the process of acquiring an MSc, and as a side effect been at least a year older, if not wiser.

A few years and many more mistakes down the line, I now teach on a programme whose main mission is to prepare students to be researchers in human and applied physiology. Practicals far outweigh the lecture time and intake is severely limited (despite the economics), so everyone's hands can get dirty. Experiments, not demonstrations, are the order of the day. Furthermore, research projects without at least the intention of publication are dirty words!

Sitting on the other side of the selection-panel desk, I see a succession of bright, talented and enthusiastic souls attempting to explain why they — holding excellent grades from a course with lecture theatres full to the brim, but having never touched

any of the relevant equipment — want to do this MSc and then go on to do a PhD.

A high proportion of those admitted onto the course have gone on to do admirably well in their subsequent PhD studies, and many have become independent and eminent researchers within their respective fields.

David Andrew Green Division of Applied Biomedical Research, Department of Physiology, King's College London, 4.19 Shepherd's House, Guys Campus, London SE1 1UL, UK e-mail: david.a.green@kcl.ac.uk

Will waste-energy plant be a waste of money?

SIR — In his entertaining Futures story 'The brown revolution' (*Nature* **455**, 564; 2008), Norman Spinrad waxes eloquent about turning faeces into energy. Some, however, are taking the idea more seriously. Swedish Biogas International is collaborating with Kettering University in Flint, Michigan (hometown of the US ambassador to Sweden, Michael Wood), to create a waste-energy plant that will recycle human faeces and turn them into renewable energy. The project will cost about $78 million, with $4 million coming from the Michigan Strategic Fund (see http://tinyurl.com/5ngr2c).

However, the value of this technology has been questioned. Taking human faeces, for example, a daily diet of 2,000 calories (8,372 kilojoules) produces an energy residue in the faeces of about 7% (586 kilojoules) — roughly equivalent to the amount of solar energy shining on one square metre for just over seven minutes (see B. B. Desai *Handbook of Nutrition and Diet*, Dekker, 2000).

Compare this with the energy in a litre of petrol: 32,000 kilojoules. One bowel movement yields the equivalent of 1.8% of a litre of petrol. We are not going to motor

very far on that. Neither does this calculation consider the energy it would take to convert faeces into energy. What is the net energy of the conversion? Is the energy ratio greater than unity, indicating that we are getting more energy out than we are putting in?

Faecal matter should be returned to the soil, from which the food that produced it originated. That closes the cycle, replenishes soil nutrients, and allows us to sustain our ecosystems and lives. Burning it, in whatever form, is akin to flaring off natural gas at the wellhead.

Gene Bazan PO Box 24, Lemont, Pennsylvania 16851, USA e-mail: genebazan@aol.com

Joining a trade union is best way to defend postdoc interests

SIR — Postdoctoral associations may be needed in many places (*Nature* **455,** 425–428; 2008). However, postdocs in the United Kingdom already have the protection of trade unions, if they choose to join.

The unions are formally recognized by employers and negotiate for their members in areas such as pay, conditions,

health and safety. Also, if an employer proposes a major change (such as a merger or department closure), trade unions will offer critical comment not just on the fate of their members but on how science — at the location, nationally and internationally — will be affected. In addition, the unions have considerable resources available with which they can provide advice and legal representation in employment disputes. For example, they recently won a 13% pay award for UK university academics, including postdocs.

I concede that postdoctoral voices struggle to be heard in unions. An association spanning different unions would be welcome, with useful roles in areas such as postdoc concerns, holding meetings and producing white papers of best practices or policy recommendations.

Being a trade unionist has empowered me. I am at the heart of negotiations, in my case within the UK Medical Research Council. For example, my institute is soon to be merged into the UK Centre for Medical Research and Innovation, and I gave evidence at the recent review on the impact of the new institute. I was talking about science, but I was invited because of my union. If I am made

redundant in the merger, I will get legal support to negotiate my redundancy terms. My colleagues who aren't members will not.

Oliver de Peyer Lab 209, Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK e-mail: opeyer@nimr.mrc.ac.uk

Detectors could spot plagiarism in research proposals

SIR — Your News story 'Entire-paper plagiarism caught by software' (*Nature* **455,** 715; 2008) follows other reports of systems to detect plagiarism (see M. Errami and H. Garner *Nature* **451**, 397–399; 2008, and S. L. Titus *et al. Nature* **453**, 980–982; 2008). Having all been involved in proposal evaluation, we believe the studies indicate that a text-matching analysis of research proposals could reduce plagiarism in subsequent publications.

For instance, when European Commission evaluators have met in the past to evaluate research proposals, they received printed copies which had to be returned before the panel members left, and had no computer access during deliberations. A plagiarism-detector using text-mining methods could be used instead of the current security measures. Such a system could, in principle, detect similarities to previous submissions and uncited sources using advanced document segmentation.

Only official agencies have access to confidential proposals and the funds to experiment with automated plagiarism-detectors. It is important that they should investigate these approaches to reducing the possibility of scientific misconduct.

Victor Maojo, Miguel García-Remesal, Jose Crespo Biomedical Informatics Group, Facultad de Informatica, Universidad Politecnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain e-mail: vmaojo@gmail.com

COMMENTARY

When consent gets in the way

As the prospect of personal genomes for all promises to revolutionize personal health records,
Patrick Taylor says that mandating consent does not protect privacy or ensure public benefit.

Translating genomic research into health care improvements will require linking genotypes with medical information that has long been considered private. Fortuitously, as genomics has progressed, so too have electronic medical records, including personal health records that are now an important part of the electronic medical information system[1]. Accompanying these developments, however, is an argument, advocated in the US Congress and elsewhere, that biomedical ethics requires subjecting any uses of electronic medical records to patient consent.

Although well-intentioned, such arguments spell trouble. Linked data are crucial for research and improving health-care quality. People might fear that information will be revealed or misused, but the impulse to block all access in the absence of consent is mistaken.

Two developments are especially foreboding. First, to support its new 'HealthVault' (a personal health record), Microsoft is projecting an ideology of total personal consent for e-data access[2,3]. Second, on Capitol Hill an unusual coalition including Microsoft, patient activists and gun lobbyists is seeking legislation prohibiting data access without patient consent, arguing that medical ethics — in particular, respect for autonomy — requires it[4,5]. This ethics invocation deserves attention. No ethical principle has transformed biomedicine as powerfully as autonomy. Joined with politics, ethics is potent enough to rewrite professional codes of conduct. Yet it is questionable whether consent-for-everything will promote privacy and public trust.

US epidemiologists, confirming ethical review-board data, say that current regulations — which permit some uses without consent — already obstruct research without increasing public trust[6]. Mandating consent increases the burden and biases research[7]. A study in Taiwan[8] showed that the elderly, the illiterate, people with low income and Taiwan aborigines were those most likely to refuse consent; even a high consent rate could not prevent selection bias against the very populations in which health questions are most compelling. Even if large cohorts muffle consent bias, consent issues will endanger research into rare diseases, including those that can cause childhood death.

The congressional arguments for autonomous control grow from deep divisions in public thought. A recent US survey concludes that 83% of the population trusts health-care providers to protect privacy; only 69% trusts researchers[9]. And 58% thinks privacy is inadequately protected. The public is evenly divided on whether researcher access is still problematic when nothing is disclosed to organizations making consumer or employment decisions. This ambivalence demands dialogue about public fears, alternatives to improve privacy protection and explanations as to why mandated consent would harm vital global research. Practical arguments can counter consent-for-everything legislation, using protected databases and methods of protecting privacy through calibrating disclosure and access[10]. Tests on genome databases indicate that addressing reidentifiability — linking anonymous data back to the person — will be difficult, but not impossible[11]. An ethics fiat should not foreclose fruitful, pointed discussion.

A private matter

Does protecting privacy require consent for all uses of health information? In answering, we must ask whether consent really protects privacy; then analyse consent against other ethical visions of autonomy; then, finally, see how fixating on consent affects other values. We cannot assume that all social goals will be met through a lemming-like coincidence of universal consent. We must address consent's conflicts with other values, even as we try to maximize both autonomy and social goods[12].

Failures in privacy protection illustrate that autonomy is not a protective measure. Personal data lost by some major retailers and US agencies were given voluntarily. Apparently, the problem was not consent but carelessness. Even if Microsoft's HealthVault protects vaulted data, it does not ensure long-term, downstream privacy. Nothing keeps patients from trading vital e-data for consumer trinkets, and no standards prevent 'buyers' of data from permitting profitable misuses by data-traders further downstream. Some fear that insurers and employers will coerce patients to permit access through bargaining strength. Let's add uncompelled disclosures: who hasn't at one time or another skipped scrolling through the fine print and clicked 'OK' to 'consents' they don't understand. Most hospital privacy notices are as unreadable to patients as professional medical literature[13]. Are the market's trade terms for valuable data sets any better, and web privacy policies so clear and fair? The options offered by companies must be ethical. Consent does not substitute for unambiguous guarantees that released data will remain private.

In Congress, autonomy is equated with ownership: my data are mine alone. The sole, controlling ethic is that patients, caricatured as 'consumers', have the right to hold or transfer data unfettered. This limited sense of autonomy is separate from social identity and community; "being a savvy consumer and participating in the vast engine of capitalism have become a substitute for being a citizen who participates in the public realm of democratic life"[14]. Ethics is reduced to autonomy; autonomy is reduced to naked choice; and a self-commodifying model of choice is substituted for richer visions of human nature and interdependence.

There is more to ethical decision-making than asking whether decisions are made autonomously. Do they take into account virtues, moral values and human narratives with less impoverished conceptions of human freedom? Are the choices good, and do they respect ethical obligations to others?

An enduring ethical position is that we should reciprocate in social arrangements through which we ourselves benefit, when the duties are fairly distributed across society[15]. A good example is improvement to health-care quality, for which access to all patient outcomes is critical. Risks from participation are low, and benefits to all are high. We depend on participation, and share a duty to participate in return. We cannot simply demand the benefit and decline the cost. Yet, a US congressional bill says no quality improvement without consent.

> "We must remember who gets left behind when consent is required."

Consent versus justice

Even if autonomy were simply an ownership choice, autonomy is only one principle among others, including beneficence and justice[16]. Living ethically requires considering the interests of others as well as one's own. Justice, for example, demands that we look at the effects of policies on health care, equality and other social premises for individual freedom and societal

benefit, and that we address state or social distortions in health-care access and disparities in outcomes. Justice loops back to autonomy, because it embodies a commitment to treating others as morally equivalent, the bedrock on which universal autonomy in society rests.

For sound reasons, some research, particularly that involving interventions on the body, requires consent altruistically given. But other research rightly proceeds on different ethical grounds. For good cause, where the primary potential harm is loss of privacy, ethics permits waiving consent given reasonable privacy protection.

How should we address privacy better? Researchers need to engage with the public on alternative solutions to genuine concerns, and clarify the impact of requiring individual consent for everything. They need to understand why patients in some countries are more comfortable with data access than in others. More legislation protecting against discrimination would be desirable. The recent US Genetic Information Nondiscrimination Act prohibits intentional discrimination in employment and insurance, but not other spheres of life. And it doesn't directly address commercial reidentification, or constrain government from expanding DNA databases through coercively linking sample extraction to public health, misdemeanours, violations or immigration. It should.

Governments should broaden privacy protections to extend across all organizations and agencies that hold sensitive information, including web service providers, pharmaceutical companies, corporate data-miners, providers of personal health records, universities and government. Reidentifiability must be addressed and prevented in cases in which extensive linkage between health and genetic information is maintained. Researchers can do

this through the design of research and database access, and governments can confer with the public on appropriate regulations.

Universities, working with government and corporate leaders, could create new research options, such as stewarded models under which independent third parties hold identifying linkages, and run queries that have passed independent ethical scrutiny. Health research and quality improvement so conducted should be exempt from redundant research reviews.

Finally, all proponents and operators of personal health records need to create standards that ensure the options they offer patients for third-party data access are ethical. Some governments require insurers to include certain policy terms and prohibit others; the result is consumer choice among alternatives that protect the insured from overreaching. Personal health records need such terms.

The price of autonomy

If we protect privacy effectively, we will not reduce ethics to autonomy, and autonomy to data ownership. Reducing ethics to ownership comes at a high price: ethics that care only about ownership and consented transfers are, by exclusion, indifferent to distributional justice and optimizing social outcomes. We must remember who gets left behind when consent is required. Genetic information must link to the full range of environmentally affected expression that diverse medical records reflect; genomics cannot exclude factors faced by indigent or illiterate people, or others who are most challenging to consent. Nor should researchers forego the studies necessary to ensure that such people also benefit from the genomic revolution. But we should protect their privacy.

This is a global-justice issue. It would be

unjust for US citizens to rest research advances on the backs of other countries' residents, while hoarding private data and reaping the benefits of global participation. And if the United States exports this self-serving autonomy, it will seriously impair global, collaborative biomedical research.

Protection of privacy is critical, but consent alone is the wrong means to protect it. Working with the public, we must preserve and explore ethical alternatives.   ■

Patrick Taylor is a lecturer at Harvard Medical School and deputy general counsel at Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.
e-mail: patrick.taylor@childrens.harvard.edu

1. Mandl, K. D. & Kohane, I. S. *N. Engl. J. Med.* **358,** 1732–1737 (2008).
2. Manos, D. Microsoft joins lawmakers, activists to demand patient privacy rights (Healthcare IT News, 18 October 2007).
3. https://account.healthvault.com/help.aspx?topicid=PrivacyPolicy
4. www.patientprivacyrights.org/site/DocServer/Letter_to_Congress_Final_10_17.07.pdf?docID=2281
5. www.patientprivacyrights.org/site/DocServer/PeelcommentsNCVHSdraftOct07REV.doc?docID=2321
6. Ness, R. B. *J. Am. Med. Assoc.* **298,** 2164–2170 (2007).
7. Ingelfinger, J. R. & Drazen, J. M. *N. Engl. J. Med.* **350,** 1452–1453 (2004).
8. Huang, N., Shih, S.-F., Chang, H.-Y. & Chou, Y.-J. *BMC Health Serv. Res.* **7,** 18 (2007).
9. Westin, A. F. *How the Public Views Privacy and Health Research* (Institute of Medicine, 2008).
10. Lowrance, W. W. & Collins, F. S. *Science* **317,** 600–602 (2007).
11. Homer, N. *et al. PLoS Genet.* **4,** e1000167 (2008).
12. Gostin, L. O. & Hodge, J. G. *Minn. Law Rev.* **86,** 1439–1479 (2002).
13. Breese, P. & Burman, W. *J. Am. Med. Assoc.* **293,** 1593–1594 (2005).
14. Churchill, L. *Daedalus* **128,** 253–274 (1999).
15. Jonas, H. *Daedalus* **98,** 219–247 (1969).
16. Beauchamp, T. L. & Childress, J. F. *Principles of Biomedical Ethics* (Oxford Univ. Press, New York, 2001).

COMMENTARY

Misdirected precaution

Personal-genome tests are blurring the boundary between experts and lay people. **Barbara Prainsack, Jenny Reardon** and a team of international collaborators urge regulators to rethink outdated models of regulation.

Since the introduction of direct-to-consumer, whole-genome testing in 2007, a handful of new companies have simultaneously fascinated and exasperated observers. 23andMe in Mountain View, California, deCODE Genetics of Reykjavik, Iceland, and Navigenics in Redwood Shores, California, are some of the companies offering consumers disease-risk information based on a genome-wide analysis. To do this, the companies look at up to a million of the single-point genetic variations known as single nucleotide polymorphisms (SNPs).

Despite the wonder of having one's genetic information probed and the allure of celebrity spit parties — publicity events that had the rich and famous providing DNA samples for analysis — these companies' business model raised hackles and fears in the research and public-health communities. Many said that clinical utility was unclear, doctors would be unable to interpret the information, customers would be unnecessarily frightened or erroneously relieved about disease risk, and privacy would be endangered in unprecedented ways[1-3].

The companies have successfully navigated legal challenges from California health authorities about lab certification and licensing, but commentators have called for regulatory oversight, or even tight regulation, of personal-genomics services[4,5]. We believe that anticipatory governance is premature without a better understanding of how SNP-based whole-genome information is used by, and what it means to, a wide range of users. At present, the only anecdotal evidence available is from wealthy (and presumably also healthy) early participants. An understanding of what a broader range of users hope to learn from this type of whole-genome information, and whether it would lead to actual life and behaviour changes, would help in assessing whether personal-genomics services are likely to be adopted in large numbers. This could happen in their current form as stand-alone services, or whole-genome data could enter clinical practice as part of patients' electronic health records, together with family histories and lifestyle information.

Personal-genomics services should not be allowed to circumvent governance for the reasons they propose. Companies argue that individuals should have the right to decide whether to take genetic tests and participate in genetic research, and that state protection is paternalistic and patronizing, impeding individual rights to consume and participate in the production of genomic knowledge. Although we welcome a shift from genetic protectionism to a situation in which individuals become experts on, and active governors of, their genomes, society should not succumb to fantasies about 'empowered' individuals making free, informed choices in an unregulated genomic marketplace.

Protectionism and empowerment are simply different sides of the same governance coin. Both imagine that good governance derives from decisions that are uninfluenced by political and economic forces. But we do not live in a world where such imaginative fictions of freedom are helpful — the close relationships between modes of producing knowledge and producing economic value are too obvious.

Converging roles

If anything, personal genomics has rendered this relationship even closer. 23andMe, for example, encourages customers to upload health, physical and lifestyle information, and to participate in genetic research. For the first time, we see research participants paying to be enrolled. Here, the notions of donor, customer, patient and activist are merging. Increasing efforts of public and non-governmental agencies and private companies — such as the National Human Genome Research Institute of the US National Institutes of Health based in Bethesda, Maryland, the National Geographic Society in Washington DC and 23andMe — to include 'people' in the design and regulation of research promise a 'democratization' of scientific practice. But democratization should not be assumed to be an *a priori* good; it can produce unexpected costs for researchers and research subjects alike[6].

The emphasis on individual empowerment often disguises the fact that personal genomics is pushing the individualization of responsibility for health one step further. The quantity of information that individuals need to consider when making choices about their health is on the rise[7]. Apart from increasing the burden of individual responsibility (and the blame for poor health), it is questionable how free and independent individual choice in this context is: although personal-genomics companies proffer education, those who sell products are seldom the best educators of their potential customers.

Regulation will be effective only if it is informed by the results of a systematic examination of these issues. We recommend that public authorities make it a priority to fund empirical research exploring what individuals expect from personal genomics, and in what way genetic-susceptibility information is likely to affect practices and lifestyle choices. The Coriell Personalized Medicine Collaborative of the non-profit Coriell Institute for Medical Research based in Camden, New Jersey, and a public–private partnership between Scripps Translational Science Institute in San Diego, California, with Navigenics and Microsoft have led by example as they explore systematically how users are making sense of personal genetic tests. Depending on the outcomes of studies such as these, governments should decide to what extent existing regulations of DNA testing should be extended to personal-genomics services, and in what contexts new legislation is necessary.

The best solution is unlikely to be the simple extension of existing regulation of laboratory tests, and of genetic testing for medical purposes[5]. This is because existing regulatory regimes of traditional medical genetic testing are based on assumptions that are no longer tenable in the post-genomics era. For example, the California Department of Public Health, when sending cease-and-desist letters to several personal-genomics companies, assumed that a medical test is a distinct entity governed by a clearly discernible set of experts: doctors and public-health authorities. This no longer holds true. Genomics blurs the boundaries that make such clear distinctions possible. A genome scan reveals information that is medical, genealogical and recreational. And those who scan and interpret the data are not distinct bodies of experts, but instead, novel configurations of geneticists, customers, ethicists, bioinformatics experts and new media executives.

Moreover, some commentators argue that the principle of medical genetic testing in a clinical context doesn't apply, because many doctors know less about genomics than personal-genomics customers themselves[8]. The

> **"Regulatory frameworks from the genetics age are ill-suited for personal genomics."**
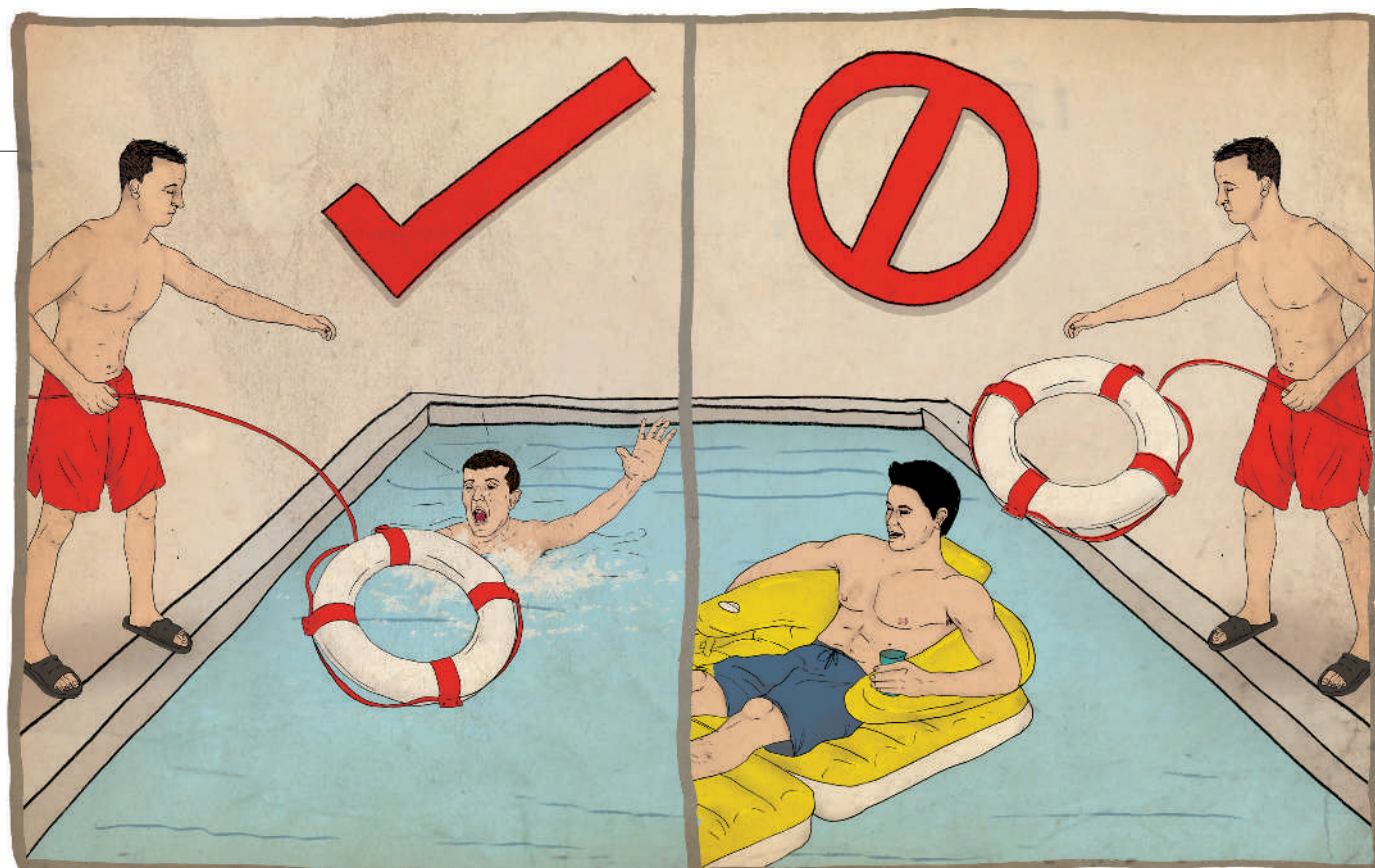
J. TAYLOR

industry tends to agree with this criticism, but perhaps for a different reason: the more that physicians need to be involved and trained, the slower the growth of the industry.

In this world of converging roles, both protectionist regulation and notions of consumer empowerment will fail, because they rely on clear boundaries that no longer hold. These are boundaries between experts and lay people; between academic knowledge and economic power; and between patients and donors. Effective responses to this situation require clarification of the novel issues created by the convergence of information about health, consumer and lifestyle choices, and genealogy; novel relations between geneticists, patients, consumers and corporate executives; and the continued intensification of collaboration, on both the research and the patient/consumer sides.

The spell is breaking

Efforts to regulate personal genomics using strategies from the genetics era miss two crucial points: the business is still very much 'in the making', embedded in dense relations between data, services, economic models and research endeavours; and it is also likely that genetic discrimination will cease to be the main concern. These two points are connected. Personal-genomics customers are already going through a process of disenchantment: it is increasingly clear how little power SNP-based readouts of a person's 'genotype' offer for predicting future ailments in an individual. Reported frustrations of 'early adopters' with the kind of information

they've received show that the fascination may be fading. Similarly, we predict that insurance companies will find little to gain from SNP data alone. SNP data are meaningful when embedded in lifestyle data, medical records and family disease histories, and this is exactly where the field will develop. Google Health (https://www.google.com/health), a free electronic health-record feature launched earlier this year, already encourages users to store medical records and family histories on the Internet. Given that Google and 23andMe are technologically and financially linked, a possible way of making use of personal-genomics test results could be to link them with other data in one's electronic health records.

The questionable predictive medical value of SNP-based testing, and enthusiastic rhetoric about empowering individuals, should not lead to the conclusion that the field should remain unregulated. However, regulatory frameworks from the genetics age are ill-suited to the task, and premature regulation could have unintended negative effects. Research needs to address the question of how people will use such data. Current arguments for regulation of this nascent industry place a premium on genetic information as a determinant of future health or illness. This is misguided, partly because the arguments rest on a distinction — that might be obliterated — between genetic data and other types of data. We should enter these waters with our eyes open, but not be afraid to get wet. ∎

Barbara Prainsack is at the Centre for Biomedicine & Society, and in the Department of Twin Research & Genetic Epidemiology at King's College London, London WC2R 2LS, UK.
e-mail: barbara.prainsack@kcl.ac.uk
Jenny Reardon is in the Department of Sociology, University of California, Santa Cruz, California 95064, USA.
Richard Hindmarsh is at the Griffith School of Environment, and Centre for Governance and Public Policy, Griffith University, 170 Kessels Road, Nathan, 4111 Queensland, Australia.
Herbert Gottweis and **Ursula Naue** are at the Life Science Governance Research Platform, Department of Political Science, University of Vienna, Vienna A-1010, Austria.
Jeantine E. Lunshof is in the Department of Social Medicine, Maastricht University, 6211 LK Maastricht, the Netherlands; and the Department of Molecular Cell Physiology, VU University Amsterdam, the Netherlands.

1. Hunter, D. J., Khoury, M. J. & Drazen, J. M. *N. Engl. J. Med.* **358,** 105–107 (2008).
2. Janssens, A. C., Gwinn, M., Bradley, L. A., Oostra, B. A., van Duijn, C. M. & Khoury, M. J. *Am. J. Hum. Genet.* **82,** 593–599 (2008).
3. Van Ommen, G. B. & Cornel, M. C. *Eur. J. Hum. Genet.* **16,** 403–404 (2008).
4. Borry, P. *Nature Biotechnol.* **26,** 736–737 (2008).
5. Gniady, J. A. *Fordham Law Rev.* **76,** 2429–2475 (2008).
6. Reardon, J. *BioSocieties* **2,** 239–256 (2007).
7. Rose, N. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century* (Princeton Univ. Press, 2006).
8. *Nature Biotechnol.* **26,** 716 (2008).

BOOKS & ARTS

Insects of war, terror and torture

Whether natural or intentional, the security threats posed by arthropods — from assassin bugs to disease-carrying pests — should be of concern to us all, explains **Kenneth J. Linthicum**.

Six-Legged Soldiers: Using Insects as Weapons of War
by Jeffrey A. Lockwood
Oxford University Press: 2008. 400 pp.
£14.99, $27.95 (hbk)

From plagues to malaria transmission, insects and other arthropods have threatened military and civilian populations throughout human history. The success or failure of military campaigns has frequently been determined by correctly anticipating the risks of diseases borne by insects and other vectors, and then mitigating against them. Recognizing this, the world's armed forces employ a large cadre of scientists with expertise in entomology or preventive medicine.

Six-Legged Soldiers describes many potential or actual uses of insects as offensive weapons during the past 100,000 years, with an emphasis on the past 300 years. Entomologist Jeffrey Lockwood describes how stinging and highly toxic insects and other arthropods have been used to cause pain and suffering to foes — from the use of bees and hornets by early humans to attack enemies, to the assassin bugs used by an Uzbek emir for torture in the early 1800s.

It is often difficult to determine whether an insect-borne threat is a natural occurrence or an intentional act. As an example, Lockwood explains how six of the ten plagues that struck Egypt, as described in the Old Testament book of Exodus, may have been caused by natural phenomena involving insects. As natural vectors of disease, insects affected many wars in recorded history, including Napoleon's campaigns, the American Civil War — in which two-thirds of the 500,000 soldiers who died were killed by diseases such as malaria and yellow fever — and the First World War.

In the Second World War, insects were developed as biological weapons; the infamous Japanese Unit 731 programme had plans to produce 5 billion plague-infected fleas per year. During the cold war there was an unprecedented level of research and development into using insects as biological warfare agents. Lockwood discusses accusations and activities concerning Korea, Vietnam, Cuba, the Soviet Union and the United States. He ends with a look to the future uses of insects in warfare,

Tiny terrorists: the assassin bug (above) and the Colorado potato beetle, or 'Amikäfer' (left), touted as a US cold-war weapon in 1950s East Germany.

including, potentially, agroterrorism, bioterrorism, insects as sentinels and detectors, and insect cyborgs.

Biological warfare is typically developed as clandestine operations. Although it may be used in propaganda campaigns to create fear among the enemy, it is poorly documented. The secret nature of this morally repugnant form of warfare is maintained to eliminate evidence that could be used by prosecutors in future international war-crimes tribunals. Lockwood relies on personal interviews and declassified and previously published documents, and he presents a wide array of accounts. He is carefully circumspect, realizing that some of the accounts may be untrue or partially true, and he qualifies his statements accordingly.

Lockwood takes care to describe accurately the scientific nomenclature of the insect taxa he is discussing, whether it be the mosquito vector of dengue fever, *Aedes aegypti*; a putative tick vector of haemorrhagic fever in the family Ixodidae; or the Mediterranean fruitfly *Ceratitis capitata*.

Six-Legged Soldiers highlights the vulnerability of the United States and other Western nations to terrorist attacks. It draws from the 1999 introduction of West Nile virus into the United States, where the disease, of unknown origin, spread from New York to California in five years. A potentially greater threat is posed to human and animal health by Rift Valley fever, another mosquito-borne disease of sub-Saharan Africa. Lockwood states that "the prognosis for curtailing Rift Valley fever by suppressing its vectors is poor", and implies that US public-health and agricultural communities are not addressing the threat. However, he fails to recognize the efforts that are underway. Outbreaks in Africa are being predicted by scientists at the US Department of Defense, NASA and the US Department of Agriculture, allowing international bodies and individual nations to enhance global vigilance. US federal, state and local agencies are developing research

agendas and formulating control strategies for vectors of Rift Valley fever.

Lockwood describes a history of collaborations in the United States between the defence department and the Department of Agriculture to develop insect-based biological weapons extending back to the Second World War. Yet he does not mention other significant collaborative efforts to protect military and civilian populations from insect bites and disease transmission, such as the development in the late 1940s of the most effective and widely used insect repellent, DEET, and the Deployed War-Fighter Protection (DWFP) programme started in 2004 to produce new insect repellents and control products and technologies to protect deployed troops. The DWFP programme has produced more than 60 peer-reviewed scientific publications including the application of RNA interference technology to potentially develop a new class of insecticide that is safe to non-targeted species. Given the paucity of effective vector-borne disease mitigation tools, the products developed in the DWFP programme will directly reduce disease.

Six-Legged Soldiers is an excellent account of the effect that arthropod-borne diseases have had on warfare. The discussions of how we are prepared, or not, for future threats from military operations, accidental introductions or bioterrorist events are pessimistic. The book highlights the need for further research to prevent, detect and mitigate vector-borne disease introductions, and to prevent globalization of entomological threats. This book will inspire readers to understand those threats and prepare new methods to combat them. ■

Kenneth J. Linthicum is director of the USDA Agricultural Research Service, Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, Florida 32608, USA. e-mail: kenneth.linthicum@ars.usda.gov

Tapping out a message

Vibrational Communication in Animals
by Peggy S. M. Hill
Harvard University Press: 2008. 272 pp.
$39.95, £25.95, €30.00

Coding and Redundancy: Man-Made and Animal-Evolved Signals
by Jack P. Hailman
Harvard University Press: 2008. 272 pp.
$39.95, £25.95, €30.00

In animal-communication research, the understanding of group behaviour is important. The development of the framework for communication networks 15 years ago has provided the field with a great conceptual advance. It takes into account that many signalling interactions do not only involve a sender and a receiver — bystanders may also eavesdrop to gain valuable information about the relative strength, aggressiveness or levels of cooperation in potential opponents or partners. Consequently, signallers may adjust their behaviour to address eavesdroppers as well as the main recipient. Such audience effects can increase levels of both aggression and cooperation in communication networks, which are seen in many diverse species across a wide range of taxa.

This framework concept links to the field of animal cognition. Animals in a group must keep track of relationships between group members to form the most beneficial coalitions, but the complexity of following these relationships increases exponentially with group size. Baboon females, for example, know both the relative rank and the matrilineal membership of all other group females. In humans, cooperation between individuals in a large group may yield benefits through indirect reciprocity — eavesdroppers are more willing to help individuals who have contributed to the public good. Two new books remind us that the physical aspects of animal communication are also important.

In *Vibrational Communication in Animals*, animal behaviourist Peggy Hill provides an up-to-date overview of this field. Because the field of vibrational communication deals with a communication channel that is alien to our own species, research can be both frustrating and exciting. Many case studies in the book read like lawsuits in which a combination of indices provides a compelling case in the absence of more direct evidence. That may be caused by the complex technical equipment required to measure the propagation of signals in the material being vibrated. But the rest of Hill's work is a beautiful case of integrative biology, highlighting anatomical and neurophysiological studies that describe the organs that receive and emit signals, and the behavioural studies needed to document that a species uses vibrational information in its communication. Owing to the introduced dichotomy between vibrational and auditory communication, scientists must exclude the auditory route as the primary information channel to conclude that animals use vibrational signals.

Hill makes a strong case that vibrational communication is widespread in animals. She uses an impressive collection of examples drawn across taxonomic groups. Particularly enjoyable is the case of the banner-tailed kangaroo rat — individuals develop their own signature foot-drumming, which they keep for life unless a new similarly drumming neighbour warrants adjustments to guarantee individual recognition. Another amazing story is about treehoppers, in which kin groups of these plant-eating insects use vibrational signals to coordinate their movements from a depleted resource to a better one — a wonderful example of groups acting as information centres.

A take-home message of Hill's book is that there are many unresolved questions that warrant more research. Signals could be varied to test if they still convey meaning, or to show that encoded information is simple. A new framework might predict under which circumstances vibrational communication will be selected over other means. A better understanding may also yield practical benefits: there are many anecdotes about certain animal species that can sense earthquakes or tsunamis and take evasive action. Overall, the book demonstrates beautifully the strength of research on animal behaviour, the appreciation of the great diversity of species and their adaptations to their specific ecological niches.

In *Coding and Redundancy*, zoologist Jack Hailman classifies man-made and animal-evolved signals according to the information coded within them. Key attributes include the type of information — binary, multivalued or multivariate — and the level of redundancy. Hailman's approach is novel, and his writing is easy to follow. Because the goal of the book is to classify, it does not say much about recent studies of animal communication. Instead,

Rat-a-tat: banner-tailed kangaroo rats drum to their own beat.

M. MCDONALD/NATUREPICTURE LIBRARY

it offers a historic background, describing how, until the late 1970s, communication was just one part of a broad natural-history data set collected by ethologists. The book ends abruptly without drawing major conclusions. The author states in the introduction that "characterising how signals encode information is only the first step in understanding animal communication". Let us hope that someone will be challenged to take the next step and combine the classification with a functional approach that considers how manipulation, deception and eavesdropping by potential friends and opponents, including predators, may have selected for specific information encoding and levels of redundancy.

These two books will provide behavioural ecologists with new ideas about the mechanisms underlying communication, which may give fresh insights into signal evolution. One can ask, as in *Vibrational Communication in Animals*, how signals might be designed and adjusted to deceive, to keep communication private, or to address an audience as well as a receiver. This interaction between proximate and ultimate questions is where we can achieve major advances in our understanding. ■

Redouan Bshary is professor of behavioural ecology at the University of Neuchâtel, Institute of Zoology, CH-2009 Neuchâtel, Switzerland.
e-mail: redouan.bshary@unine.ch

History out of the ether

Blessed Days of Anaesthesia: How Anaesthetics Changed the World
by Stephanie J. Snow
Oxford University Press: 2008. 256 pp.
£16.99, $34.95

I am no hero. All my life I have tried to avoid situations that might produce physical danger or pain; my dentist is now well trained. In my lectures on pain and anaesthesia, I advised medical students to choose their anaesthetist more carefully than their surgeon. Without good anaesthesia, surgeons could achieve few of their triumphs.

Thus I would seem an ideal reader for Stephanie Snow's new book, *Blessed Days of Anaesthesia: How Anaesthetics Changed the World*. But it leaves me dissatisfied. Its chapter titles are promising — 'Women, Sex, and Suffering', 'On Battlefields', 'The Dark Side of Chloroform'. Snow's knowledge of nineteenth-century medicine and society is considerable. Yet the book lacks a central concept.

Former Australian prime minister Paul Keating said that "a soufflé does not rise twice", and, compared with its predecessor *Operations Without Pain: The Practice and Science of Anaesthesia in Victorian Britain* (Palgrave Macmillan, 2006), Snow's new book is thin, despite its grander title. *Blessed Days of Anaesthesia* is also more hubristic because it directs its attention mainly towards British practice and society. Naturally the United States and the ether story must be mentioned, as should nineteenth-century war, especially because the American Civil War had great effects on medicine and wider society. Snow deals with these to some extent, but glimpses beyond her homeland are rare.

This bias is odd because the story of anaesthesia is global. In quoting from a crucial letter from the Boston botanist Jacob Bigelow to his London colleague Francis Boott in December 1846 — with its detailed account of the famous operations under general anaesthesia in the Massachusetts General Hospital — Snow omits a prophetic sentence: "This is something which will go round the world". And it did. Gwen Wilson's magisterial *One Grand Chain: The History of Anaesthesia in Australia 1846–1962* (Australian and New Zealand College of Anaesthetists, 1995) gives a splendid account of the way anaesthesia spread, notably through the maiden voyage of the *Pekin,* which departed Southampton in February 1847 for Ceylon, and the impressive work of its surgeon, Thomas Bell, seemingly with a procedure in every port.

Snow's new book is not a real medical history, nor is it seriously concerned with medicine or society beyond England and Scotland. But it seeks to link developments in anaesthesia with changing social, philosophical, scientific and religious attitudes in those countries.

She begins by setting her account within the context of that metamorphosing national culture, but history is presented as a series of brief anecdotes. The cases presumably humanize an abstract story and show that medicine is a personal matter for patients, their families and their doctors. The book opens with the overused drama of the novelist Fanny Burney's horribly painful mastectomy in 1811, the same story with which Thomas Dormandy began his account of general anaesthesia in *The Worst of Evils: The Fight Against Pain* (Yale University Press, 2006). Snow's approach has benefits but tends to diffuse the story and obscure her inchoate structure. The effect is like the recall of one's visits to great galleries through a collection of souvenir postcards: the experience is inevitably diminished.

Snow is at her most compelling when describing the criminal use of chloroform. The tale of the death of the sexually ambiguous London grocer Edwin Bartlett in 1886, the murder trial of his wife, Adelaide, and her still-contested acquittal is engrossing. It reminds us that the rapid take-up of new technology is not always benevolent, as with modern cases of drug-facilitated 'date rape'.

The author is rightly proud of the pioneering work of her husband's ancestor, John Snow — anaesthetist and epidemiologist — of whom she wrote in her previous book, "For Snow, experimental science was the anchor and mainstay of medicine". Even so, we would now describe those early anaesthetists as cavalier, experimenting on themselves around their dinner tables and rushing to use successful agents on their patients. We have far greater concerns for safety today, but not much more sense of what anaesthesia really is, because an understanding of the true nature of consciousness remains intractable. The question of serendipity versus planning in medical discovery, the need to be alert to the unexpected, the ineluctable impetus to improve the lot of patients: these are some of the tensions and fascination of medical history. Anaesthetists are, accordingly, the intellectuals of clinical medicine and obligatory pragmatists.

Blessed Days of Anaesthesia peters out in a chapter that, too hurriedly and trivially, tries to tie up anaesthetic developments in the twentieth century. This story requires a book of its own, not simply because in that century anaesthesia became far safer and protean in its scope, but because modern respiratory physiology is the powerful achievement of anaesthetists.

> **"We have far greater concerns for safety today, but not much more sense of what anaesthesia really is."**

Snow concludes with an undistinguished poem by an American physician. I recognize her point: when medicine becomes a metaphor for art we know that it has achieved a profound importance. However, I prefer to think of T. S. Eliot's image: "the evening is spread out against the sky; like a patient etherised upon a table". That is art and medicine as one, a thrilling way in which anaesthesia has changed an unexpected world. ■

John Carmody is a medical scientist at the University of Sydney, NSW 2006, Australia.
e-mail: jcarmody@med.usyd.edu.au

Oppenheimer (Gerald Finley, centre) realizes the power of his bomb in *Doctor Atomic*.

Q&A: Opera for the end of the world

The dawn of the nuclear era finds its voice in *Doctor Atomic*, an opera about J. Robert Oppenheimer and the making of the first atom bomb. With a new production showing in New York, composer **John Adams** explains how physicists have reacted to the work, and how writing it has changed his view of nuclear weapons.

What is the setting for the opera?
It mainly takes place during the night leading up to the detonation of the first atomic bomb, code-named the Trinity test, in New Mexico on the morning of 16 July 1945. Just as the plutonium sphere had been winched up on a tower over the desert, an electrical storm blew in, causing huge anxiety. There was pressure to test the bomb as soon as possible because Russia wanted a piece of Japan.

How is the story told?
Peter Sellars compiled a one-of-a-kind libretto using historical sources for every line of sung text. Some of Oppenheimer's words are drawn from a top-secret memorandum that discussed target choices. Because he was a cultured person, we used his favourite poets for moments of lyrical flight or hallucination. Exhausted and nervous, with a dawning awareness of the horror of his creation, our Oppenheimer sings from Charles Baudelaire's poetry, the sacred text of the Bhagavadgita and the John Donne sonnet from which he supposedly drew the name 'Trinity'.

Did Oppenheimer face opposition about dropping the bomb?
After two years of utter focus on the engineering feat, the war in Europe was suddenly over and there were rumours that the bomb would be dropped on civilians. A petition [to warn the Japanese] was circulated by young scientists who naively thought it would end up on the President's desk. But people have different recollections. After one rehearsal, an 80-year-old physicist who had worked at the Los Alamos National Laboratory in New Mexico came up to me and said, "I want you to know that there wasn't a single person who wasn't happy as hell that we dropped that bomb on Japan."

How has working on the opera changed your view of nuclear weapons?
I've been thinking about the use of the bomb in Japan for eight years now and I still can't tell you whether I think it was the right decision. We were facing a land invasion where a million people could have been killed. If the bomb had not been used in Japan, I'm almost certain it would have been used in the Korean war a few years later. It's just human nature.

Have you received any criticism from scientists?
The first words sung by the chorus used to be "matter can be neither created nor destroyed but only altered in form." Marvin Cohen, a professor of physics at the University of California, Berkeley, wrote to me to say that's not strictly right [because a fission bomb turns matter into energy]. I tried to fix it at the dress rehearsal of the San Francisco production in 2005 but the chorus panicked. I have since corrected it.

The opera mentions physicist Edward Teller's concerns about the bomb igniting the air around it. Why did you include this?
Enrico Fermi had voiced his concern that the bomb might cause an atmospheric meltdown, and Teller once calculated the odds of this. By 1945, that possibility was not considered seriously, but I wanted to keep the discussion in the opera because this weapon was a tipping point in the relationship between our species and the planet. Starting on that morning, we had it in our power to destroy the world.
Interview by **Jascha Hoffman**, a writer based in New York.
e-mail: jascha@jaschahoffman.com

Doctor Atomic
The Metropolitan Opera, New York City
Until 13 November 2008.
Broadcast live from New York to cinemas worldwide on 8 November 2008.
English National Opera, London
Opens on 25 February 2009.

ESSAY

Language: a social history of words

Language evolved as part of a uniquely human group of traits, the interdependence of which calls for an integrated approach to the study of brain function, argue **Eörs Szathmáry** and **Szabolcs Számadó**.

Our ability to communicate using language is often cited as the element that sets us apart from other animals. Although language is not uniquely human in all aspects — dogs and apes, for example, can learn the meaning of many words — it almost certainly merits special status. This is because, more than any other attribute, language was probably key to the development of the set of traits that makes humans unique.

The evolution of language probably occurred in concert with the evolution of many of the other traits we associate with being human, such as the ability to fashion tools or a strong propensity to learn. If this is true, it suggests that we shouldn't be trying to understand one characteristically human trait in isolation from the others. Moreover, instead of the brain being a collection of separate modules, each dedicated to a specific trait or capacity, humans are likely to have a complex cognitive architecture that is highly interconnected on multiple levels.

Enhanced communication would have aided humans at least as far back as the Late Pleistocene, around 120,000 years ago. By this point, humans were proficient at hunting large game. Indeed, the advantages that groups of hunters would have derived from better communication may have helped drive the evolution of language at first. But language was almost certainly later co-opted for a wide array of activities. The diversity of behaviours that appeared during the Late Pleistocene, including fishing, use of pigments, and tool and weapon making, as well as the rate at which they emerged, suggest that by the time humans acquired the full set, they could also communicate using complex language.

Many of these developments had a clear social context: making spear points or using pigments, for example, must have relied on learning from other group members. Studies of chimpanzees show that without language, the spread of knowledge in basic tool-using tasks, such as using a stone hammer and anvil to crack a nut, is highly inefficient.

In fact, the bulk of our grammatical machinery enables us to engage in the kinds of social interaction on which the efficient spread of these tasks would have depended. We can combine sentences about who did what to whom, who is going to do what to whom, and so on, in a fast, fluent and largely unconscious way. This supports the notion that language evolved in a highly social, potentially cooperative context, involving and requiring at least three attributes: shared attention, shared intentionality and theory of mind. In other words, individuals would have been able to pay attention to the same scene or object as others; be aware that they must act as a group in order to achieve a common goal; and attribute mental states to others as well as to themselves.

Uniquely human

The probable emergence of modern language in the context of these other capacities points to the evolution of a uniquely human set of traits. We've barely begun to probe the architecture of this 'suite', but there is little to suggest that each capacity evolved one by one, or that they could be lost independently without harming at least some other traits in the set.

Take cooperation. In humans, practices such as staying faithful to one sexual partner and sharing food suppress competition within groups. These can be upheld more easily with language, because language means details can be agreed on and conflicts cleared up. Hunting in packs is more efficient if hunts can be planned and plans communicated. And both cooperation and communication using language are easier if people can pay attention to the same thing, are aware that others have states of mind that may differ from their own, and realize that they need to act as a group.

Moreover, some of the traits in the suite require very similar types of operation. Language is not critical for making tools; the steps involved can be spread by non-verbal teaching and imitation, or learnt through individual experience. But, in the same way as syntax, the 'action grammar' of complex manipulations involves hierarchical processing. When we fashion a tool, just as when we form a sentence, we construct it from simpler units.

Joined-up development

Evidence supporting the close-knit evolution of traits comes, for example, from experiments showing that people who struggle with grammar also have difficulties drawing hierarchical structures, such as a layered arrangement of matches.
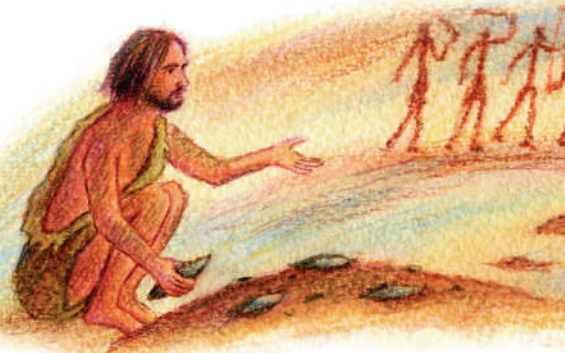
In addition, recordings of brain activity suggest that the same cognitive structures are involved in linguistic processing and tool making. In a recent study, a group of people was asked to make a specific type of ancient stone axe, which required different types of work to be done in a specific order. Brain images taken during the process revealed activation in a region in the right hemisphere. This is analogous to a region in the left hemisphere called Broca's area that is involved in language. The right-hemisphere area is also known to take on language-processing duties when the left hemisphere is damaged at an early age.

Establishing how the genes underpinning the various traits interact may likewise provide support for the idea that the human traits are closely interrelated. Of course, genes don't code for language or the capacity to fashion tools. They code for proteins and RNA molecules that serve structural, functional and regulatory roles. Take the *FOXP2* gene. When mutated, this disrupts motor control of the mouth and face, and the shaping of words, such as regular verbs in the English past tense. *FOXP2* is expressed in vertebrates other than humans and in human tissues other than the brain. In birds and mammals, it seems to be involved in the general development of neural circuitry that ensures the smooth, fast delivery of sequential movements.

That the genes involved in a cognitive trait affect other traits, and have effects that interact with each other, is business as usual for complex behaviour. But the result is likely to be a network of interacting effects, in which evolution in one trait builds on an attribute already

> **"Cultural evolution has shown us that one word can be worth a thousand genes."**

G. BECKER

modified as a by-product of selection acting on another. The nature of the gene networks underpinning complex behaviour suggests that several genes will have been selected for because they enhanced proficiency in a range of tasks — whether in social, linguistic or tool-use domains.

Analysing whether the genes involved in, say, cooperation, influence other traits in the suite is an exciting avenue for research. As a first step, it would be useful to clarify the functions of the hormones oxytocin and arginine vasopressin. Certain genetic variants of these hormones' receptors have been linked to autism, a brain disorder that impairs social interaction by disrupting language development and the capacity to pay attention to the same thing as other people. Genetic changes in the vasopressin receptor gene also correlate with how people allocate funds to other players in a game of experimental economics investigating altruism.

Cutting out the knife

The functional interdependence of characteristically human cognitive traits, plus the interlinked gene networks likely to underpin them, point to a complex cognitive architecture. The distinct gene networks and brain regions underpinning each trait can be likened to the separate towers of a castle, which are connected by common rooms and corridors. This picture could potentially replace the much-used 'Swiss army knife' view of the brain. Long advocated in evolutionary psychology, this proposes that separate cognitive modules perform specific functions. Several observations that are at odds with the knife model could be explained by the more holistic castle view.

For example, as shown by people with syntax deficiencies being poor at drawing hierarchical structures, capacities can be synergistic, where proficiency in one domain means proficiency in another. In addition, disruption in a specific element of one trait is often accompanied by a problem in another capacity. For instance, people who have trouble formulating grammatical sentences tend to fare worse than average in IQ tests because of poor short-term memory. This is consistent with the view that genes affecting a combination of cognitive capacities are far more common than genes whose disruption would harm a single trait.

The disorder known as specific language impairment also poses problems for the Swiss army knife view. As its name suggests, this condition is generally considered to affect only language. Nonverbal IQ is apparently left largely intact. But, although in the 'normal' range, children with this syndrome tend to show significantly lower IQ scores than their siblings. And even adult sufferers often have problems in capacities aside from language, for example, in auditory processing and motor skills.

Together, these observations suggest that if the modular, Swiss army knife picture of the brain is applicable at all, it may be so only to the final outcome of development. Associations of specific brain regions with certain traits are clearly evident, but these should be assessed at different stages in development and investigated as part of a multilayered network of interactions. A more holistic approach is likely to reveal 'intermediate capacities' that have emerged as a result of evolutionary selection acting on multiple traits. Analogical reasoning — the ability to transfer information from one object to another and deduce something about the second object from the first — may fall into this category, as this is critical in tool use and tool making, but probably also opened up possibilities for complex language.

The evidence strongly suggests that language evolved into its modern form embedded in a group of synergistic traits. However, language almost certainly holds special status over the other traits in the set. More than any other attribute, language is likely to have played a key role in driving genetic and cultural human evolution.

Language enables us to pass on cultural information more efficiently than can any other species. It's taken about 40 million years, for example, for five agricultural systems to appear in fungus-growing ants. Human agriculture diversified on a massive scale in just a few thousand years. Language makes it easier for people to live in large groups and helps drive cumulative cultural evolution — the build-up of complex belief systems, and the establishment of laws and theories over several generations. It has allowed us to construct a highly altered social and physical world, which has in turn shaped our evolution. Cultural evolution has shown us that one word can be worth a thousand genes. Language was the key evolutionary innovation because it built on important cognitive prerequisites and opened the door to so much else. ∎

Eörs Szathmáry and **Szabolcs Számadó** are at the Biological Institute of Eötvös Loránd University, 1/c Pázmány Péter sétány, H-1117 Budapest, Hungary. E.S. is also at the Collegium Budapest and the Parmenides Center for the Study of Thinking, Munich, Germany.
e-mail: szathmary@colbud.hu

See http://tinyurl.com/6hxb56 for further reading. For more on Being Human, see www.nature.com/nature/focus/beinghuman.

NEWS & VIEWS

E. LESLIE

POPULATION BIOLOGY

Case of the absent lemmings

Tim Coulson and Aurelio Malo

Changing weather patterns, producing the wrong kind of snow, have transformed the population dynamics of lemmings in northern Scandinavia. The knock-on effects have been felt throughout the ecosystem.

A colleague from Oslo once told me that when the Bible was translated into Norwegian, mention of plagues of locusts was replaced with plagues of lemmings. The logic behind this change was that most Norwegians knew nothing of locusts, but were all too familiar with periodic explosions in lemming numbers. The story is apocryphal, with references to lemmings only scrawled by the translator in the margin. Yet these scribbles suggest that lemming outbreaks have been a feature of northern ecosystems for the past millennium. But now the outbreaks, at least in some areas, have stopped. On page 93 of this issue, Kausrud *et al.*[1] explore the underlying reasons.

Norway lemmings (*Lemmus lemmus*) are remarkable animals. These rodents can live for three or four years, spending their winters beneath the snow and feeding mostly on moss. A female can produce up to three litters a year, with as many as 12 young per litter. Lemmings occasionally become super-abundant when large numbers of young survive[2]. In northern Norway in 1970, lemmings were so common that snowploughs were used to clear the vast numbers of squashed animals from roads. Outbreaks don't last long: food becomes scarce, and lemmings will then often disperse en masse in search of greener pastures. On occasion, desperate to find food, they jump into water and start swimming. This behaviour led to the myth that lemmings commit suicide.

In northern Scandinavia, lemming outbreaks typically occur once every three to five years. Or they used to. In the past 15 years, localized outbreaks have either stopped or occur less frequently[3]. The cause of this change is the subject of debate, partly because the reason that rodent populations often show periodic outbreaks is itself controversial[4–8]. Fluctuating predation, food availability or quality, and climate variability have all been proposed as plausible mechanisms generating these population cycles. Whatever the cause, it is clear that in parts of northern Europe something is now preventing these rodents from periodically producing large numbers of surviving young[3].

Kausrud *et al.*[1] analyse a 27-year time series of lemming numbers from one site in Norway. They first demonstrate statistically that climate change means that Norway now gets a lot of the wrong sort of snow. Lemmings do well when warmth from the ground melts a small layer of snow above it, leaving a gap between ground and snow. This subnivean space provides warmth and allows lemmings to feed in relative safety from many of the animals that eat them. Climate change now means that the subnivean space does not exist for as much of each year as it used to. Worse still, the space itself is less likely to form: warmer temperatures mean that snow melts and refreezes, producing a sheet of ice that prevents lemmings from feeding on the moss.

The wrong sort of snow therefore means that food is hard to come by, keeping warm is challenging, and being eaten is more likely. Kausrud *et al.* use their statistical associations to construct a predictive model of lemming dynamics. This model, fitted to data from before the outbreaks stopped, predicts the observed cessation, providing compelling evidence that changing snow conditions are a major factor in the change in lemming population dynamics.

The researchers then go on to show that the reduction in the frequency of lemming outbreaks has knock-on consequences for the wider ecosystem. They argue that the scarcity of lemmings means that predators such as foxes turn their attention to other species, including willow grouse and ptarmigan, adversely affecting their populations. Evidence for changes in the numbers of species other than lemmings in these northern ecosystems is convincing. But although the mechanism that Kausrud *et al.*[1] propose — a shift in predation patterns — is plausible, it is speculative.

The critical reader will complain that the story is based on correlations. Although this is true, it is often the only way to study populations and the consequences of changing climate for ecosystems[9]. The collection of detailed long-term data on the dynamics of free-living populations of animals and plants rarely attracts the same excitement as

genomics or particle physics, yet such data are vital in characterizing the consequences of climate change for the natural world on which we depend. Describing and predicting such effects of climate change will help us prepare for, and possibly minimize, adverse affects. Kausrud et al.[1] elegantly show the value of detailed long-term ecological data, and demonstrate the benefits of maintaining existing studies and instigating new ones.

By the time the Norwegian translator of the Bible got to the book of Revelations, he had stopped making references to lemmings, so we do not know whether the cessation of outbreaks foretells the imminent arrival of the four horsemen of the apocalypse. However, we do now understand that climate change has made lemming outbreaks much less common, which has in turn affected the fragile ecology of northern ecosystems. This research[1] provides

a striking example of how climate change can modify the workings of the natural world — raising the question of what consequences such change might have for us. ∎

Tim Coulson and Aurelio Malo are in the Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, Berkshire SL5 7PY, UK.
e-mail: t.coulson@imperial.ac.uk

1. Kausrud, K. L. et al. Nature 456, 93–97 (2008).
2. Turchin, P. et al. Nature 405, 562–565 (2000).
3. Ims, R. A., Henden, J.-A. & Killengreen, S. T. Trends Ecol. Evol. 23, 79–86 (2008).
4. Aars, J. & Ims, R. A. Ecology 83, 3449–3456 (2002).
5. Lambin, X. & Krebs, C. J. Oikos 61, 126–132 (1991).
6. Angerbjorn, A., Tannerfeldt, M. & Erlinge, S. J. Anim. Ecol. 68, 34–49 (1999).
7. Seldal, T., Andersen, K. J. & Hogstedt, G. Oikos 70, 3–11 (1994).
8. Hanski, I. et al. Ecology 82, 1505–1520 (2001).
9. Coulson, T. et al. Science 292, 1528–1531 (2001).

ASTROPHYSICS

An illuminating dark halo

Stéphane Colombi

A large simulation reveals that most of the detectable signal from dark matter in our Milky Way probably comes from the main, smooth Galactic halo, rather than from small clumps.

Most of the mass of the Universe is believed to be in the form of dark matter — an invisible component that has so far been only indirectly detected through the effects of its gravity on visible matter. In the theory of supersymmetry in particle physics, there is a corresponding dark-matter-particle candidate that interacts only very weakly with the rest of the Universe, and is thus very difficult to detect directly. There is, however, a general feeling in the astronomical community that the search for dark matter is now at a turning point. This feeling stems from the recent start of the largest particle accelerator in the world (the Large Hadron Collider), which could provide clues about the nature of dark matter, and from the advent of high-energy astrophysics observations, such as γ-ray observations carried out by NASA's Fermi Gamma-ray Space Telescope. Such observations are potentially able to detect dark-matter particles indirectly through their annihilation radiation. On page 73 of this issue, Springel et al.[1] show that the primary and probably most easily observable annihilation signal is produced by the diffuse dark-matter component rather than the very small clumps in the main halo of our Galaxy (Fig. 1).

The challenges in determining the nature of dark matter are not only experimental. At a time when observations are about to start providing data, it is necessary to understand in detail how dark matter is distributed in our neighbourhood, in particular in the halo

surrounding our Galaxy (an extended, ellipsoid-shaped dark-matter structure), in order to make predictions about the expected annihilation signal. During the past few years, there has been controversy about the nature of the clustering of dark matter inside galactic haloes, and particularly the mass and distance from

our Solar System of the dark-matter structures that are likely to contribute to the annihilation signal that could be measured by the high-energy astrophysics experiments.

There are two ways to predict the properties of dark matter inside galactic haloes. The first involves simplifying the geometry of the problem, and making predictions using relatively simple but robust analytical calculations[2]. Although these calculations are rigorous and free from any numerical artefacts, the oversimplification of the geometry can lead to questionable results. The second approach, used by Springel et al.[1], involves performing sophisticated numerical experiments on supercomputers.

Springel et al. study the dynamics of dark matter in a cosmological background — the expanding Universe — by modelling the dark-matter distribution with a set of macroparticles that interact which each other only through gravitational forces. Each macroparticle represents a huge number of actual dark-matter particles. Because the gravitational force is very long-range in nature, the authors simulate a large volume of the Universe and zoom in on a region where a halo similar to that of our Galaxy is formed. In that smaller region, the resolution of the simulation is increased, enabling many macroparticles of smaller mass to trace all the fine details of dark-matter dynamics.

Springel and colleagues' simulations are developed in the framework of the cold dark matter (CDM) hypothesis, which is now the commonly accepted model for the formation of large-scale structures in the Universe. One of the hurdles to performing simulations in CDM models is achieving numerical convergence at small scales, or equivalently at small masses. Within the CDM hypothesis, the consensus is that the smallest dark-matter structures formed

V. SPRINGEL

Figure 1 | Dark matter around the Milky Way. The Galactic dark-matter halo contains a remarkable hierarchy of structures of different sizes. But according to Springel et al.[1], it is the diffuse, smooth component of dark matter in the halo that is likely to dominate the expected annihilation radiation of the corresponding dark-matter particles. (More pictures and movies are available at www.mpa-garching.mpg.de/aquarius.)

have sizes comparable to that of the Solar System and masses equivalent to that of Earth[3]. These small structures (substructures) would then have merged together to form larger ones, and so on, forming a full hierarchy of structures within structures. The largest structures correspond to haloes of rich clusters of galaxies.

So the question is whether or not dark-matter simulations have enough resolution to resolve the smallest structures. The bigger the number of dark-matter particles used in the simulations, the larger the number of sub-structures detected. But at what stage can we be sure that numerical convergence is achieved?

Springel et al. answer this question to a large extent[4] by identifying and tracing dark-matter structures and substructures in a very robust way. To achieve that end, they perform several simulations with various resolutions — that is, with different numbers of particles, but with the same initial configuration. They are then able to cross-identify the substructures found in the different simulations and perform a quantitative, unprecedented convergence study of the fine details in the distribution of dark matter in our Galactic halo.

They conclude that, in fact, the main contribution for indirect dark-matter detection should come from the smooth component of the halo of our Galaxy instead of its substructures, at variance with some earlier analyses[5,6]. If these results are confirmed, astronomers should take them into account in future analyses of γ-ray observations, particularly when trying to disentangle the contribution of dark matter from that of other γ-ray sources, such as those found in the plane of our Milky Way. The debate about the nature and small-scale distribution of dark matter remains open. At the very least, however, Springel and colleagues have made a great advance in the field of computational cosmology. ■

Stéphane Colombi is at the Institut d'Astrophysique de Paris, CNRS UMR 7095/Université Pierre et Marie Curie, 98 bis boulevard Arago, 75014 Paris, France. e-mail: colombi@iap.fr

1. Springel, V. et al. Nature 456, 73–76 (2008).
2. Mohayaee, R. & Salati, P. Preprint at http://lanl.arxiv.org/abs/0801.3271 (2008).
3. Diemand, J., Moore, B. & Stadel, J. Nature 433, 389–391 (2005).
4. Springel, V. et al. Preprint at http://lanl.arxiv.org/abs/0809.0898 (2008).
5. Calcáneo-Roldán, C. & Moore, B. Phys. Rev. D 62, 123005 (2000).
6. Berezinsky, V., Dokuchaev, V. & Eroshenko, Y. Phys. Rev. D 68, 103003 (2003).

BIOCHEMISTRY

Enzymes under the nanoscope

Anthony J. Kirby and Florian Hollfelder

Small-scale interactions of substrates with an enzyme's active site — over distances smaller than the length of a chemical bond — can make big differences to the enzyme's catalytic efficiency.

When Richard Feynman died in 1988, he left behind the following words on his blackboard: "What I cannot create, I do not understand." His message certainly resonates with protein engineers. When it comes to making enzymes, we are clearly missing something, because artificial enzymes cannot yet be designed that match natural catalysts in efficiency. Reporting in the *Journal of the American Chemical Society*, Sigala et al.[1] explain at least part of the reason why this is so. Tiny variations (on the scale of 10 picometres, where 1 picometre is 10^{-12} metres) in the binding interactions and molecular packing in an enzyme's active site can make a remarkable difference to the efficiency of enzymatic catalysis.

Any biochemistry textbook will tell you that enzymes catalyse reactions by binding their substrates' transition states — high-energy arrangements of atoms that form during reactions — more tightly than the ground states. This differential recognition lowers the energy barrier for reaction, and usually occurs because the transition state fits better into the active site than does the ground state, and/or because the active site stabilizes any charges in the transition state more than those in the ground state[2].

Strong support for this idea comes from transition-state analogues (TSAs) — stable molecules designed to mimic the shapes and charges of transition states. TSAs are highly efficient inhibitors of enzyme catalysis because their tight binding to, and slow release from, enzymes' active sites blocks the turnover of native reactions[3]. Such molecules can even be used as templates to generate antibodies. Because these antibodies bind to TSAs, they should also bind to transition states for reactions modelled by the TSAs, thus catalysing those reactions. Such 'catalytic antibodies'[4] are arguably the best models of enzymes that we have, but the reaction-rate accelerations of these proteins are still tens of billions of times smaller than those of many enzymes[5,6].

Available tools for protein engineering clearly lack the subtle touch that is required to prepare effective designer enzymes. For example, site-directed mutagenesis (a method in which specific amino acids in proteins are replaced with others) is commonly used to investigate the roles of individual amino acids in catalysis. But the sizes of naturally occurring amino acids vary by discrete increments of at least one chemical-bond length (roughly 140 picometres), whereas breaking bonds in a transition state extend by only about 20 picometres, compared with the same bonds in the ground state. The modifications that we can make to active sites are therefore larger in scale than those that enzymes have evolved to detect. Further complications arise because, in a highly interconnected protein structure, a single amino-acid change introduced by site-directed mutagenesis can create all sorts of structural changes elsewhere in the enzyme.

Sigala et al.[1] now report an approach for identifying the distance scale at which enzymes recognize the structural reorganization of substrates during reactions. Because transition states are, by definition, short-lived high-energy species that are not amenable to direct analysis, Sigala et al.[1] had to investigate the effects on enzyme binding of structural variations in a TSA. They did this using a battery of modern analytical techniques — including high-resolution X-ray crystallography, nuclear magnetic resonance spectroscopy, quantum-mechanical calculations and TSA-binding measurements — that allowed them to resolve a difficult problem with unprecedented precision.

The enzyme chosen for study was keto-steroid isomerase (KSI), which catalyses the migration of a carbon–carbon double bond in a wide variety of ketosteroid substrates, by way of a negatively charged 'dienolate' intermediate (Fig. 1a, overleaf). The intermediate, and thus the transition state that leads to it, is stabilized by hydrogen bonding to hydroxyl (OH) groups in the side chains of two amino acids in the active site. These groups constitute an 'oxyanion hole' — a region of hydrogen-bonding groups capable of accommodating and stabilizing the negative charge that develops in the dienolate. Such oxyanion holes are held firmly in position by tight packing of local hydrophobic residues.

The hydrogen bonds that stabilize the transition states in KSI also bind substrates in their ground states, but are presumed to 'tighten up' as the reaction proceeds. Sigala et al. monitored this tightening process using negatively charged phenolate ions as probes (Fig. 1b). Phenolates have a similar geometry and charge distribution to that of the dienolate, and bind to the active site of KSI using the same hydrogen bonds[7]. In KSI substrates, a carbon–oxygen double bond lengthens as the transition state forms, and the negative charge on the oxygen increases. Similarly, the length of an analogous carbon–oxygen bond in phenolates can be varied by changing a substituent on the phenolate; the electron density
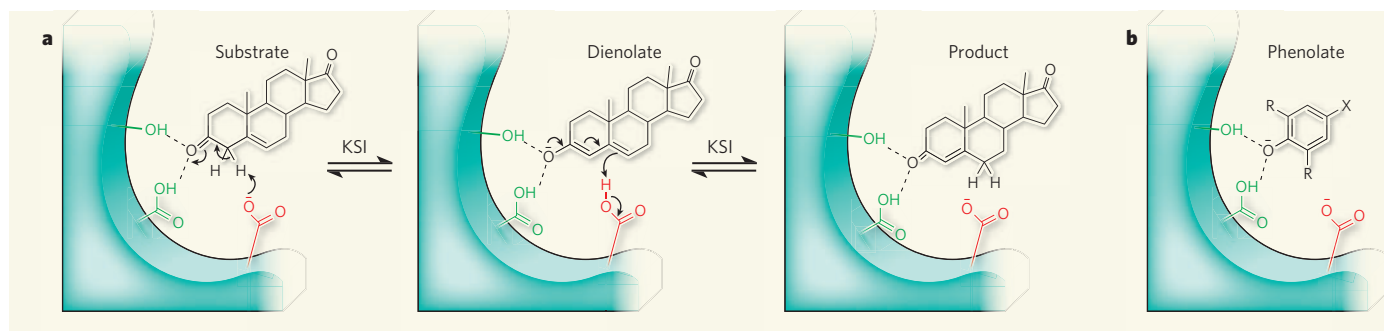
Figure 1 | Enzyme-catalysed isomerization. The enzyme ketosteroid isomerase (KSI) catalyses a reaction in which a carbon–carbon double bond in the substrate moves to a new position in the molecule. **a**, The side chain of an amino acid (red) in the active site triggers the reaction, which proceeds through a dienolate intermediate. Other side chains (green) stabilize the dienolate and the transition state that leads to it by forming hydrogen bonds (dashed lines) to its fully or partly negatively charged oxygen. These hydrogen bonds also bind the substrate and the product, albeit more weakly. Curly arrows indicate electron movement during the reaction. **b**, Sigala *et al.*[1] investigate the control of KSI over the position of transition states during reactions, using phenolate ions as mimics of dienolates. Both the length of the carbon–oxygen bond and the electron density on the oxygen depend on the substituent X and on the bulkiness of the substituents R (R can be either hydrogen or fluorine).

on the oxygen changes at the same time.

According to the accepted mechanism for KSI-catalysed reactions, increasing the electron density on the oxygen of a phenolate should strengthen (and shorten) the hydrogen bonds that bind the molecule to the active site, and so reinforce binding to the enzyme. Sigala *et al.*[1] observe that this is indeed the case, but find that the pattern is disrupted if the phenolates are made slightly bulkier. When the hydrogen atoms attached to the carbons on either side of the oxygen are replaced by fluorine atoms (which are marginally larger and have higher electron density than hydrogens), increasing the electron density on the oxygen makes binding of the phenolate to the active site weaker, even though the hydrogen bonds should have been strengthened. This could be because the fluorine atoms start to clash (either electrostatically or physically) with the groups of the oxyanion hole as the hydrogen bonds try to become tighter, and thus shorter.

The crucial finding is that shortening of the hydrogen bonds by as little as 10 picometres is prevented by forces in and around the oxyanion hole, suggesting that the level of control exerted by the active site on the positions of its substrates operates on this stringently small scale. This result has wide-reaching implications: it defines experimentally the distance scale on which enzymes can distinguish geometric rearrangements of atoms, and determines the energetic consequences of this constraint. The picometre-precision of KSI also explains why protein engineering to produce enzymes that have new or altered functions has proved so difficult.

Sigala and colleagues' work brings one particular set of experimental tools to bear on a complex problem of fundamental importance, and will certainly concentrate the minds of those in the field. The same issue can and will be approached in other ways, and might well provide a range of answers that are specific to the system under investigation. We look forward to the development of a consensus. It will be interesting to see how other molecular probes can be used to map out the furnishings of active sites and to define and compare the distance scales for catalysis. Meanwhile, the belief that electrostatic and geometric complementarity of active sites and transition states is central to enzyme catalysis has become better defined. And, to accept Feynman's implicit challenge, what we understand, we might one day be able to create. ∎

Anthony J. Kirby is in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. Florian Hollfelder is in the Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK.
e-mails: ajk1@cam.ac.uk; fh111@cam.ac.uk

1. Sigala, P. A. *et al. J. Am. Chem. Soc.* **130,** 13696–13708 (2008).
2. Pauling, L. *Chem. Eng. News* **24,** 1375–1377 (1946).
3. Wolfenden, R. *Biophys. Chem.* **105,** 559–572 (2003).
4. Mader, M. M. & Bartlett, P. A. *Chem. Rev.* **97,** 1281–1302 (1997).
5. Hilvert, D. *Annu. Rev. Biochem.* **69,** 751–793 (2000).
6. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. *Nature* **383,** 60–63 (1996).
7. Kraut, D. A. *et al. PLoS Biol.* **4,** e99 (2006).

GLOBAL CHANGE

Climate's astronomical sensors

Michel Crucifix

A re-evaluation of the relationship between Earth's orbital parameters, ice-sheet extent and ocean circulation sets further puzzles for those trying to disentangle cause from effect in long-term climatic changes.

Earth's climate 'feels' the slow changes in the parameters of our orbit around the Sun. The great ice sheets of the Northern Hemisphere are one sensor, in that they are sensitive to the amount of solar energy they receive in summer. Lisiecki *et al.*[1] (page 85 of this issue) provide evidence that ocean dynamics also responds to orbital changes, and not just in the north.

Much of our life is controlled by the rhythms of days and seasons — not surprisingly, given that the Sun is our ultimate source of energy. Earth's atmosphere senses the rhythms of days and seasons, too, but both atmosphere and oceans may respond to the much longer astronomical cycles that affect incoming solar radiation. In 1976, Hays *et al.*[2] described how they tackled this problem. They collected deep-sea sediments in the Southern Ocean, dated them according to depth, and analysed the oxygen-isotope composition of the calcium carbonate remains of foraminifera preserved in the sediments. This quantity is a proxy for ice-age conditions: isotopic composition indicates whether climate at any time was glacial — with large ice sheets in the Northern Hemisphere and low temperatures in the deep oceans — or interglacial, as today. Hays *et al.* then plotted this measure against time to estimate the frequency spectrum. Several of the dominant glacial oscillation periods they found corresponded perfectly to the astronomical periods calculated analytically by Berger[3]: 19,000 and 23,000 years for climatic precession; 41,000 years for changes in obliquity.

So, what are precession and obliquity? Earth revolves around the Sun following an elliptic figure. The climatic-precession parameter tells us what time in the year we reach perihelion — that is, the closest point to the Sun, when Earth is globally exposed to the maximum amount of incoming solar radiation. Perihelion is

presently reached on 3 January; it will be reached in July in 11,000 years and again in January in 22,000 years. Obliquity is the angle between the Equator and Earth's orbital plane. Changes in this angle are responsible for the seasons: the larger it is, the more energy the polar areas receive in summer. Neither precession nor obliquity modifies the total amount of energy reaching Earth in one full year. Only eccentricity — the orbital deviation from the circular — does that, but the effect is so small that it is neglected in most theories. Eccentricity does, however, modulate the amplitude of the effect of precession with periods of 100,000 and 400,000 years[3].

Given that astronomical cycles hardly modify the global amount of incoming solar energy, the climate's astronomical sensors must be sensitive to the seasonal and spatial distribution of this energy. In that respect, the response of ice sheets immediately comes to mind. The amount of ice melting every year depends on the amount of solar energy absorbed during the warm season; the total ice mass is therefore expected to decrease when obliquity is high and perihelion is reached around summer. As early as 1876, John Murphy suggested that summer insolation could control glacial cycles[4]. History records the name of Milutin Milankovitch[5] as the father of this theory, however, because of the firm mathematical foundation he provided for it (although he missed some crucial aspects of the ice sheets' response[6]).

But candidate astronomical sensors other than ice sheets have been proposed, most notably in two papers[7,8] published as part of the SPECMAP project, which aims to rationalize the chronology represented by different palaeoclimate records. These papers downgraded the Milankovitch mechanism to a second-order effect, and attributed the prime cause of glacial–interglacial cycles to the response of Arctic sea ice to northern summer insolation. This Arctic response would have led to the development of northern ice sheets through a somewhat convoluted causal pathway involving circulation in the North Atlantic Ocean and changes in the concentration of atmospheric carbon dioxide.

Several drawbacks have been identified in the SPECMAP model[9], but Lisiecki et al.[1] have delivered the coup de grâce. They began by noting that SPECMAP was not supported by good palaeoenvironmental records of the
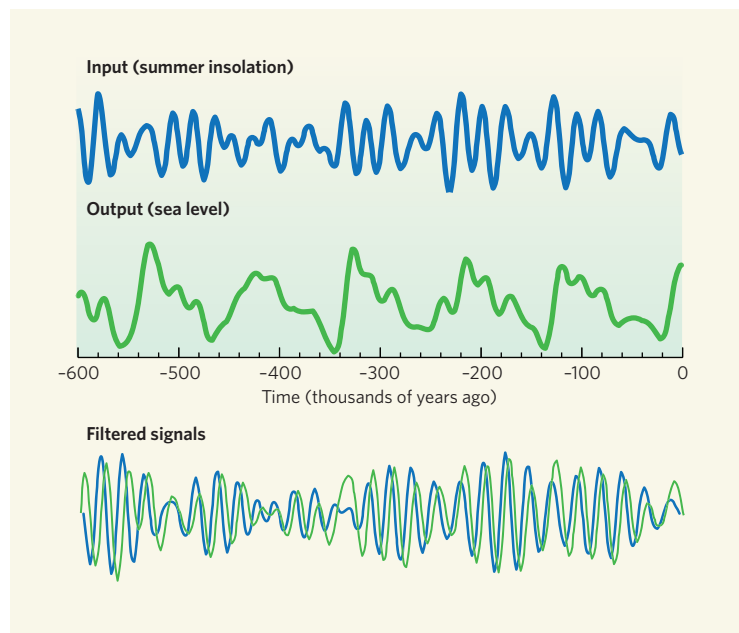
Figure 1 | Illustration of the linear signal analysis used by Lisiecki et al.[1]. This example considers the output of a simple dynamical system (green, simulated sea-level[11]) forced by a known input (blue, summer insolation in the Northern Hemisphere[12]). Both signals are filtered to extract their variance in a given frequency band (here, around 21,000 years, which corresponds to Earth's climatic precession). It is then verified that their phases, estimated by means of a Hilbert transform, are coherently related to each other. Such is the case here, with output lagging input by 1,500 to 6,500 years (90% confidence). This procedure confirms that the input effectively controls the system. But it does not guarantee that it is the cause of the large cycles in the output signal. In the artificial case tested here, these cycles are known to be autonomous: they would occur even without external forcing. The forcing simply acts as a clock, which has the effect of improving output predictability. Likewise, Lisiecki et al.[1] show that obliquity and precession control ocean circulation, but not the extent to which glacial cycles depend on this external forcing.

deep-ocean circulation. They instead used 30 archives of a naturally occurring isotopic indicator (the isotopic ratio of carbon in foraminifera shells) known to be sensitive to the distribution of water masses in the ocean. The archives are sufficiently broadly distributed geographically to provide a good idea of the global ocean circulation dynamics over the past 250,000 years.

Lisiecki et al. then essentially replicated the SPECMAP analysis procedure: band-pass filtering of time-series data to isolate the fraction of the signal thought to respond to precession and obliquity, and then assessing how this signal lagged the orbital elements (Fig. 1). The surprising result is that, when obliquity is high, the Atlantic Ocean tends to be dominated by deep water of Nordic origin — the opposite of the SPECMAP prediction. Moreover, when Earth is near its perihelion at the time of summer in the Northern Hemisphere, the Atlantic seems to be dominated by water of southern rather than northern origin.

With this, the view that the Arctic is the main 'front-end' orbital sensor of the climate system becomes hard to defend: the two orbital configurations (high obliquity and Northern Hemisphere perihelion) have the similar effect of increasing summer insolation in

the Arctic. How could they lead to opposite ocean responses? Lisiecki et al.[1] remark that things would be easier to explain if the ocean responded to summer insolation of the Southern Ocean, but we are left with conjectures to explain the mechanism. This is a challenge for those running general circulation models of the ocean and atmosphere.

Finally, another challenge merits mention. Lisiecki et al. used linear time-series analysis techniques to decipher the influence of orbital elements on climate. This is perhaps good enough to point out first-order effects, but climate is a nonlinear system. For example, it took about 100,000 years to build the big ice sheets that existed on the Earth of our mammoth-chasing ancestors, but those ice sheets largely disappeared within 10,000 years. That's not typical of a linear system. In fact, we are still unsure that orbital variations are necessary to explain glacial–interglacial cycles[10,11].

We need a more systematic way of developing and applying nonlinear statistical models to test our understanding of the slow dynamics of climate. The task is not straightforward — how, for instance, do we rigorously account for dating uncertainties in sediments without becoming trapped in circular arguments (palaeoclimate scientists know this as the 'orbital tuning' problem)? Yet it is the only way to answer the crucial question of how far ahead glacial–interglacial cycles can be predicted. ∎

Michel Crucifix is at the Institut d'Astronomie et de Géophysique G. Lemaître, Université catholique de Louvain, BE-1348 Louvain-la-Neuve, Belgium. e-mail: michel.crucifix@uclouvain.be

1. Lisiecki, L. E., Raymo, M. E. & Curry, W. B. *Nature* **456**, 85–88 (2008).
2. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
3. Berger, A. *Nature* **268**, 44–45 (1977).
4. Murphy, J. J. *Q. J. Geol. Soc. Lond.* **32**, 400–406 (1876).
5. Milankovitch, M. *Canon of Insolation and the Ice-age Problem* (Engl. transl. of original 1941 publ., Narodna biblioteka Srbije, Belgrade, 1998).
6. Weertman, J. *Nature* **261**, 17–20 (1976).
7. Imbrie, J. *et al. Paleoceanography* **7**, 701–738 (1992).
8. Imbrie, J. *et al. Paleoceanography* **8**, 699–735 (1993).
9. Ruddiman, W. F. *Quat. Sci. Rev.* **25**, 3092–3112 (2006).
10. Tziperman, E., Raymo, M. E., Huybers, P. & Wunsch, C. *Paleoceanography* **21**, doi:10.1029/2005PA001241 (2006).
11. Saltzman, B. & Maasch, K. A. *Trans. R. Soc. Edinb.* **81**, 315–325 (1990).
12. Berger, A. *J. Atmos. Sci.* **35**, 2362–2367 (1978).

HUMAN GENETICS

Individual genomes diversify

Samuel Levy and Robert L. Strausberg

The link between a person's genetic ancestry and the traits — including disease risk — that he or she exhibits remains elusive. Routine sequencing of the genomes of an African and an Asian individual offer a step forward.

The rapid progress in genetic screening assays and DNA sequencing techniques promises to increase our understanding of the complex relationship between the human genetic make-up (the genotype) and its associated traits (the phenotype). For example, using the composite human genome sequences[1–3], genome-wide association studies have identified regions that control specific traits through single nucleotide polymorphisms (SNPs) — the most common form of genetic variation. In this issue, Bentley et al.[4] (page 53) and Wang et al.[5] (page 60) detail the development and application of a high-throughput technology for sequencing DNA to decipher the genomes of two people, one of West African descent and the other of Han Chinese descent. This advance provides a technology that might eventually relate specific sequences and regions of DNA directly to human phenotypes.

Although genome-wide association studies can establish a link between a genetic locus marked by adjacent SNPs and its associated phenotype, they do not automatically identify the implicated nucleotide's position, as they use only a fraction of human SNPs. Genome-wide association studies were used because of their relatively low cost compared with the technological challenge and high cost of sequencing genomes in large human populations. Sequencing the genomes of many individuals would overcome the problem of identifying which nucleotide(s) are implicated in a phenotype, as long as the procedure could be performed accurately and completely. From such data sets, DNA variants can be identified, and the frequency with which they occur in humans who carry a particular trait — such as a disease — can then be compared with their frequency in people who lack that trait. Thus, all genetic variants contributing to the trait can be identified, giving a more complete picture of the biology involved.

The genomes of the anonymous African and Asian individuals supplement the existing sequenced genomes of two people of European origin, Craig Venter[6] and James Watson[7]. Both teams involved in the latest work[4,5] used the Illumina GA sequencing instrument, in which sequencing is performed by synthesizing fluorescently detectable DNA molecules, using the DNA from the genome being sequenced as a template. In a single cycle, this platform can produce more than 40 million discrete 'reads' of 35 nucleotides from either end of a 200- or a 2,000-nucleotide DNA fragment. Compared with the instruments used to complete the initial human genome sequence[1–3], the Illumina GA generates three to four orders of magnitude more sequence per operation cycle. This instrument therefore joins the 454 Life Sciences sequencer[7] as yet another 'next generation' technology for sequencing individual human genomes.

How do the two new genome sequences allow a better understanding of human genetics? Both studies[4,5] confirm that it is possible to routinely sequence the genome of an individual to discover the wide spectrum of DNA variations that it harbours. Of course, this process is greatly facilitated by having a reference human genome against which to compare sequence data from the two individuals. This allows the identification of SNPs, as well as insertion/deletion polymorphisms and structural variations (Fig. 1, overleaf). Extensive validation of the SNPs detected shows that sequencing accuracy is high. A strength of this

OPTICS

Metamaterial Persian carpets

Metamaterials gained renown as a way of creating invisibility cloaks — devices that could make an object 'disappear' before one's eyes. Less well known is that they can also act as detectors for biological compounds. Writing in *Optics Express*, Bingham et al. describe two-dimensional metamaterials designed so that, when exposed to electromagnetic radiation, their resonant frequencies coincide with those of vitamin H (C. M. Bingham *et al. Optics Express* **16,** 18565–18575; 2008). The resonant frequencies of vitamin H occur in the terahertz range, and these results thus provide an example of biodetection in that frequency regime.

The properties of metamaterials lie in their structure rather than their chemical composition. One asset of these man-made materials is that they can be engineered to possess a precise response to electromagnetic radiation. Bingham and colleagues created metamaterials with designs that mimic several types of symmetry observed in nature, using both square and hexagonal tiles. Their tiles, the unit cells of metamaterial structures (shown on left of picture), consist of up to three different subunits. The overall structures (shown on right) look rather like a Persian carpet.

To maximize the electromagnetic response of a metamaterial, the unit cells must be tightly tessellated — that is, the gaps between tiles must be minimized. But why incorporate more than one subunit into a tile? The advantage is that the metamaterial preserves the different electromagnetic properties of each subunit: a material formed with three distinct subunits is resonant at three different frequencies. A triple-resonator metamaterial allows a biological compound to be identified more accurately because there are three frequency-match points of comparison.

With this in mind, the authors simulated metamaterial structures computationally to find the best materials for the job. They then made the best designs, shone terahertz radiation on them and recorded the electromagnetic response. As predicted, metamaterials with structures that combined three distinct subunits (such as that pictured on the lower right panel) resonated at three distinct frequencies, the individual frequencies of the different subunits.

As the authors had hoped, the simulated and experimental resonances of their metamaterials were a good match for those of vitamin H. This match could therefore form the basis of a biodetector. Bingham et al. have found that their multi-subunit tiling techniques can create multi-resonator metamaterials that can be used as biodetectors. But that is not all. Their metamaterials could potentially detect hazardous chemicals.
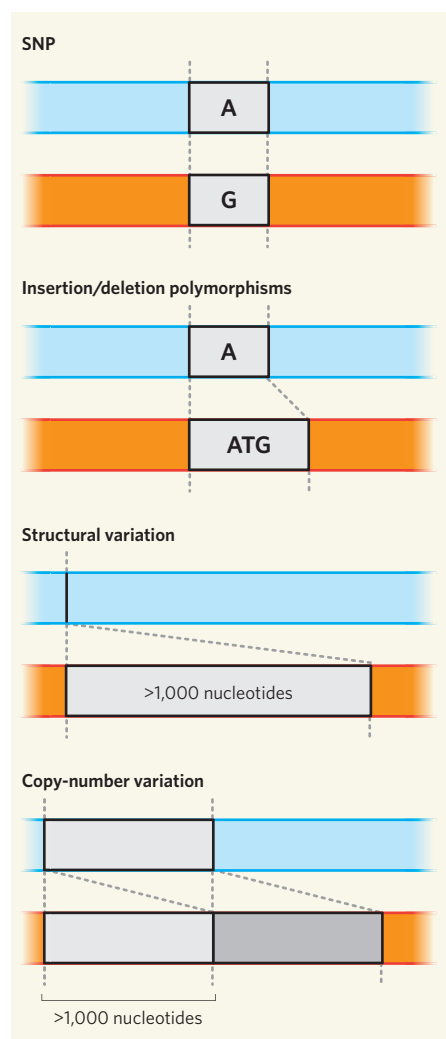
Ana Lopes

OSA

Figure 1 | Genomic variations. The latest whole-genome sequences of two humans confirm[4,5] that individual genomes vary in several respects. The types of variability in inheritance include: variations in single nucleotides (SNPs); insertion or deletion of several nucleotides; insertion or deletion of thousands of nucleotides (structural variation); and duplication or multiplication of DNA segments more than 1,000 nucleotides long (copy-number variation).

latest approach is the extent of deep sequencing achieved, which aids SNP identification.

The advantages of obtaining these two genomes, such as the identification of DNA variations, indicate that their usefulness will ultimately be much broader than simply demonstrating the technological milestone of relatively low-cost sequencing. But some goals remain. As the genomes were reconstituted on the basis of alignments with existing reference genomes, the set of non-SNP variants that are absent in the reference genome will be incomplete. For example, in these studies, the detection of structural variants — insertions or deletions of thousands of nucleotides at any one position on a chromosome — is preferential for deletions. This is because such insertions come from sequenced reads that will not overlap with the existing reference genome. There are two

possible solutions to this detection bias. One would be to sequence larger DNA fragments whose ends overlap with sequences on the reference genome[8]. Alternatively, all sequenced reads could be assembled independently, before mapping them to a reference human genome[6].

Another deficiency of the four genomes[4–7] is that they do not accurately define copy-number variants at the nucleotide level. These forms of genetic variation arise from the insertion of multiple copies of DNA segments that may include whole genes and that have been increasingly implicated in, among other disease phenotypes, neurological disorders[9,10].

Our genomes are not just collections of DNA variation: parental inheritance also dictates specific associations between neighbouring variations. Knowledge of these associations will ultimately help us discover whether and how much of an aberrant protein is produced by each of our cells and how these events contribute to observed phenotypes. The association between neighbouring variations across all 23 pairs of human chromosomes is referred to as haplotype assembly, and has not yet been completely achieved in any of the individual genomes sequenced.

These limitations notwithstanding, the approach of Bentley[4], Wang[5] and their colleagues represents a substantial advance in the sequencing of individual human genomes. Together with the other two genomes sequenced[6,7], they reinforce the catalogue of variants that exist in human genomes — SNPs in the millions, insertion/deletion polymorphisms in the hundreds of thousands and structural variants in the thousands. The numbers of these variants do not directly tell us how such polymorphisms contribute to the wide spectrum of human traits. But they do provide a necessary step towards accurately defining genomic loci that are likely to be implicated in those traits.

With such rapid advances in next-generation technologies, and with 'third generation' technologies emerging, this is just the beginning of the era of the individual genome. Soon, association studies using complete individual genomes will become the approach of choice for understanding the complexity of human biology and disease. The latest advances have broad implications for expediting that goal. ∎

Samuel Levy and Robert L. Strausberg are at the J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: slevy@jcvi.org

1. Lander, E. S. *et al. Nature* **409,** 860–921 (2001).
2. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
4. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
5. Wang, J. *et al. Nature* **456,** 60–65 (2008).
6. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
7. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
8. Kidd, J. M. *et al. Nature* **453,** 56–64 (2008).
9. Marshall, C. R. *et al. Am. J. Hum. Genet.* **82,** 477–488 (2008).
10. Walsh, T. *et al. Science* **320,** 539–543 (2008).

See Editorial, page 1.

50 YEARS AGO

Nobel Prize for Chemistry: Dr. F. Sanger, F.R.S. — The award has been made for his researches on the structure of the protein hormone insulin … When he began his investigations on insulin, Dr. Sanger first devised the use of dinitrofluorobenzene for the identification and estimation of the free amino-groups of proteins or peptides, and this method has since been widely adopted … Dr Sanger's methods and example have stimulated much research in the investigation of protein structure, the limits of which have yet to be visualized, and they make clear the possibility that insulin may be completely synthesized in the laboratory, although this is unlikely to occur for some time to come.
From *Nature* 8 November 1958.

100 YEARS AGO

Windmills and Water-Wheels. By R. S. Ball — As is natural, the author commences his book with a reference to the, said to be, not distant day when all the coal, and all the oil, in the world will have been used up, and mankind, in order to sustain itself, will have to rely wholly upon the water-wheel and the windmill for that tremendous amount of energy which will be necessary to keep the immense population of the earth in the state of comfort which it has, with the progress of civilization, attained.

ALSO:

A meeting of the Child Study Society was held on October 29, when a paper was read by Miss Alice Ravenhill on the results of an investigation into hours of sleep among elementary-school children … The evil of insufficient sleep is widespread. Parents must be roused to a sense of the importance of the subject, and the enforcement of the laws on the employment of children should be rendered obligatory upon local authorities.
From *Nature* 5 November 1908.

50 & 100 YEARS AGO

HUMAN GENETICS

Individual genomes diversify

Samuel Levy and Robert L. Strausberg

The link between a person's genetic ancestry and the traits — including disease risk — that he or she exhibits remains elusive. Routine sequencing of the genomes of an African and an Asian individual offer a step forward.

The rapid progress in genetic screening assays and DNA sequencing techniques promises to increase our understanding of the complex relationship between the human genetic make-up (the genotype) and its associated traits (the phenotype). For example, using the composite human genome sequences[1–3], genome-wide association studies have identified regions that control specific traits through single nucleotide polymorphisms (SNPs) — the most common form of genetic variation. In this issue, Bentley et al.[4] (page 53) and Wang et al.[5] (page 60) detail the development and application of a high-throughput technology for sequencing DNA to decipher the genomes of two people, one of West African descent and the other of Han Chinese descent. This advance provides a technology that might eventually relate specific sequences and regions of DNA directly to human phenotypes.

Although genome-wide association studies can establish a link between a genetic locus marked by adjacent SNPs and its associated phenotype, they do not automatically identify the implicated nucleotide's position, as they use only a fraction of human SNPs. Genome-wide association studies were used because of their relatively low cost compared with the technological challenge and high cost of sequencing genomes in large human populations. Sequencing the genomes of many individuals would overcome the problem of identifying which nucleotide(s) are implicated in a phenotype, as long as the procedure could be performed accurately and completely. From such data sets, DNA variants can be identified, and the frequency with which they occur in humans who carry a particular trait — such as a disease — can then be compared with their frequency in people who lack that trait. Thus, all genetic variants contributing to the trait can be identified, giving a more complete picture of the biology involved.

The genomes of the anonymous African and Asian individuals supplement the existing sequenced genomes of two people of European origin, Craig Venter[6] and James Watson[7]. Both teams involved in the latest work[4,5] used the Illumina GA sequencing instrument, in which sequencing is performed by synthesizing fluorescently detectable DNA molecules, using the DNA from the genome being sequenced as a template. In a single cycle, this platform can produce more than 40 million discrete 'reads' of 35 nucleotides from either end of a 200- or a 2,000-nucleotide DNA fragment. Compared with the instruments used to complete the initial human genome sequence[1–3], the Illumina GA generates three to four orders of magnitude more sequence per operation cycle. This instrument therefore joins the 454 Life Sciences sequencer[7] as yet another 'next generation' technology for sequencing individual human genomes.

How do the two new genome sequences allow a better understanding of human genetics? Both studies[4,5] confirm that it is possible to routinely sequence the genome of an individual to discover the wide spectrum of DNA variations that it harbours. Of course, this process is greatly facilitated by having a reference human genome against which to compare sequence data from the two individuals. This allows the identification of SNPs, as well as insertion/deletion polymorphisms and structural variations (Fig. 1, overleaf). Extensive validation of the SNPs detected shows that sequencing accuracy is high. A strength of this

OPTICS

Metamaterial Persian carpets

Metamaterials gained renown as a way of creating invisibility cloaks — devices that could make an object 'disappear' before one's eyes. Less well known is that they can also act as detectors for biological compounds. Writing in *Optics Express*, Bingham et al. describe two-dimensional metamaterials designed so that, when exposed to electromagnetic radiation, their resonant frequencies coincide with those of vitamin H (C. M. Bingham et al. Optics Express **16,** 18565–18575; 2008). The resonant frequencies of vitamin H occur in the terahertz range, and these results thus provide an example of biodetection in that frequency regime.

The properties of metamaterials lie in their structure rather than their chemical composition. One asset of these man-made materials is that they can be engineered to possess a precise response to electromagnetic radiation. Bingham and colleagues created metamaterials with designs that mimic several types of symmetry observed in nature, using both square and hexagonal tiles. Their tiles, the unit cells of metamaterial structures (shown on left of picture), consist of up to three different subunits. The overall structures (shown on right) look rather like a Persian carpet.

To maximize the electromagnetic response of a metamaterial, the unit cells must be tightly tessellated — that is, the gaps between tiles must be minimized. But why incorporate more than one subunit into a tile? The advantage is that the metamaterial preserves the different electromagnetic properties of each subunit: a material formed with three distinct subunits is resonant at three different frequencies. A triple-resonator metamaterial allows a biological compound to be identified more accurately because there are three frequency-match points of comparison.

With this in mind, the authors simulated metamaterial structures computationally to find the best materials for the job. They then made the best designs, shone terahertz radiation on them and recorded the electromagnetic response. As predicted, metamaterials with structures that combined three distinct subunits (such as that pictured on the lower right panel) resonated at three distinct frequencies, the individual frequencies of the different subunits.

As the authors had hoped, the simulated and experimental resonances of their metamaterials were a good match for those of vitamin H. This match could therefore form the basis of a biodetector. Bingham et al. have found that their multi-subunit tiling techniques can create multi-resonator metamaterials that can be used as biodetectors. But that is not all. Their metamaterials could potentially detect hazardous chemicals.
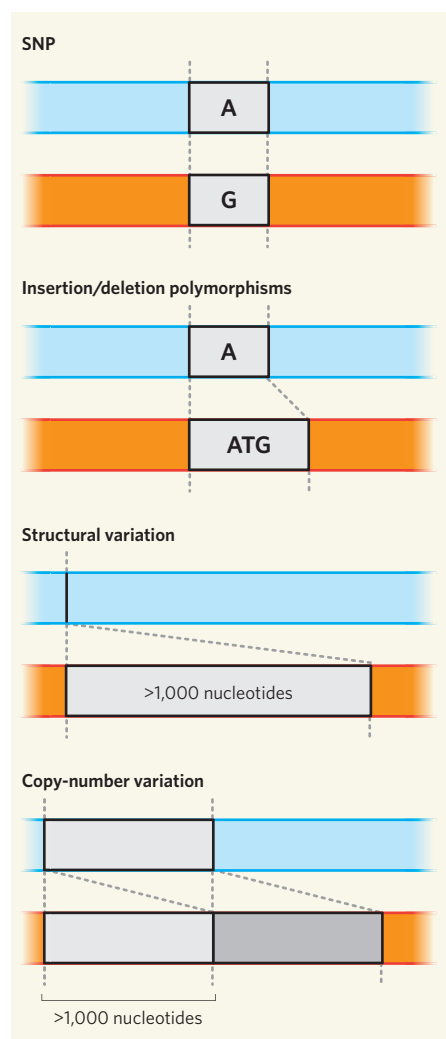
Ana Lopes

OSA

Figure 1 | Genomic variations. The latest whole-genome sequences of two humans confirm[4,5] that individual genomes vary in several respects. The types of variability in inheritance include: variations in single nucleotides (SNPs); insertion or deletion of several nucleotides; insertion or deletion of thousands of nucleotides (structural variation); and duplication or multiplication of DNA segments more than 1,000 nucleotides long (copy-number variation).

latest approach is the extent of deep sequencing achieved, which aids SNP identification.

The advantages of obtaining these two genomes, such as the identification of DNA variations, indicate that their usefulness will ultimately be much broader than simply demonstrating the technological milestone of relatively low-cost sequencing. But some goals remain. As the genomes were reconstituted on the basis of alignments with existing reference genomes, the set of non-SNP variants that are absent in the reference genome will be incomplete. For example, in these studies, the detection of structural variants — insertions or deletions of thousands of nucleotides at any one position on a chromosome — is preferential for deletions. This is because such insertions come from sequenced reads that will not overlap with the existing reference genome. There are two

possible solutions to this detection bias. One would be to sequence larger DNA fragments whose ends overlap with sequences on the reference genome[8]. Alternatively, all sequenced reads could be assembled independently, before mapping them to a reference human genome[6].

Another deficiency of the four genomes[4–7] is that they do not accurately define copy-number variants at the nucleotide level. These forms of genetic variation arise from the insertion of multiple copies of DNA segments that may include whole genes and that have been increasingly implicated in, among other disease phenotypes, neurological disorders[9,10].

Our genomes are not just collections of DNA variation: parental inheritance also dictates specific associations between neighbouring variations. Knowledge of these associations will ultimately help us discover whether and how much of an aberrant protein is produced by each of our cells and how these events contribute to observed phenotypes. The association between neighbouring variations across all 23 pairs of human chromosomes is referred to as haplotype assembly, and has not yet been completely achieved in any of the individual genomes sequenced.

These limitations notwithstanding, the approach of Bentley[4], Wang[5] and their colleagues represents a substantial advance in the sequencing of individual human genomes. Together with the other two genomes sequenced[6,7], they reinforce the catalogue of variants that exist in human genomes — SNPs in the millions, insertion/deletion polymorphisms in the hundreds of thousands and structural variants in the thousands. The numbers of these variants do not directly tell us how such polymorphisms contribute to the wide spectrum of human traits. But they do provide a necessary step towards accurately defining genomic loci that are likely to be implicated in those traits.

With such rapid advances in next-generation technologies, and with 'third generation' technologies emerging, this is just the beginning of the era of the individual genome. Soon, association studies using complete individual genomes will become the approach of choice for understanding the complexity of human biology and disease. The latest advances have broad implications for expediting that goal. ■

Samuel Levy and Robert L. Strausberg are at the J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: slevy@jcvi.org

1. Lander, E. S. et al. Nature **409,** 860–921 (2001).
2. Venter, J. C. et al. Science **291,** 1304–1351 (2001).
3. International Human Genome Sequencing Consortium Nature **431,** 931–945 (2004).
4. Bentley, D. R. et al. Nature **456,** 53–59 (2008).
5. Wang, J. et al. Nature **456,** 60–65 (2008).
6. Levy, S. et al. PLoS Biol. **5,** e254 (2007).
7. Wheeler, D. A. et al. Nature **452,** 872–876 (2008).
8. Kidd, J. M. et al. Nature **453,** 56–64 (2008).
9. Marshall, C. R. et al. Am. J. Hum. Genet. **82,** 477–488 (2008).
10. Walsh, T. et al. Science **320,** 539–543 (2008).

See Editorial, page 1.

50 & 100 YEARS AGO

OBITUARY

George Emil Palade (1912–2008)

A founding father of modern cell biology.

George Emil Palade died on 7 October at the age of 95. He was among the greatest scientists of the twentieth century, whose momentous discoveries in cell biology are still actively pursued by many laboratories worldwide.

The son of a philosophy professor and a teacher, Palade was born in Jassi (Iaşi), the former capital of Moldavia, the eastern province of Romania. He studied medicine at the University of Bucharest. Having spent the Second World War in the medical corps of the Romanian army, he moved to Istanbul shortly before moving on to New York City in 1946 for postdoctoral studies at New York University.

Following a short stint there, in a life-changing event Palade was invited by Albert Claude to join his laboratory at the Rockefeller Institute for Medical Research — now Rockefeller University. The previous year, Claude and his colleagues Keith Porter and Ernest Fullam had published the first electron micrograph of an animal cell grown in culture, describing a "lace-like cytoplasmic network", later named the endoplasmic reticulum. Furthermore, Claude and his collaborators George Hogeboom and Walter Schneider had recently developed procedures involving differential centrifugation to break up tissues and to separate cellular components into three main fractions — nuclei, mitochondria and 'microsomes'. So Palade joined an already famous laboratory that was on the cusp of even greater discoveries.

He soon became a key member of the lab, contributing vigorously to optimizing methods for both cell fractionation (such as introducing sucrose solutions for better preservation of cellular organelles) and electron microscopy (using osmium tetroxide to get better contrast). These technical advances facilitated many pivotal discoveries by Palade and his colleagues throughout the 1950s and 1960s, among them a detailed description of the membranes of mitochondria and chloroplasts. His other achievements included the discovery in 1955 of "a small particulate component of the cytoplasm" — often referred to as the 'Palade granule' until it morphed into the 'ribosome' in 1958 — and the description in 1963 with Marilyn Farquhar of "junctional complexes in various epithelia", which connect epithelial cells together.

With Claude moving back to Belgium in 1949, the partial disassembly of 'the Rockefeller group' began, ending with Porter's departure to Harvard in 1961. In fond memory of Porter's contributions, a picture of him, with the title "Our father who art at Harvard", decorated the Palade lab at Rockefeller for many years.

In parallel with his discoveries using electron microscopy, Palade sought to understand the function of these newly defined cellular structures. Biochemical

studies with Philip Siekevitz in his lab on the microsome fraction were published in classic papers in which microsomes were identified as broken and sealed bits of the endoplasmic reticulum. Subsequently, through *in vivo* labelling with ^{14}C-leucine and isolation of labelled chymotrypsin protein from cell fractions, Palade and Siekevitz showed that this protein was primarily synthesized in microsomes. These results led to the proposal that the endoplasmic reticulum is the synthesis site for secretory proteins, an idea further supported by experiments carried out by David Sabatini and Colvin Redman, who demonstrated that the initial event in the protein secretion pathway was directional release of nascent polypeptide chains into the microsomal lumen.

With another colleague, Jim Jamieson, Palade developed the technique of pulse-chase labelling in tissue slices, which allowed the pathway of secreted proteins to be tracked in time and traced within cells. One important, but initially controversial, postulate was that secretory proteins are transported in quanta — in vesicular carriers that bud from a donor membrane and deliver their contents by fusion to a target membrane.

Palade ran his laboratory very informally. There were no regular lab meetings. Instead, there were bimonthly seminars, at which lab members were introduced to ideas by speakers from other labs. Often Palade summarized the essence of a presentation, particularly if the speaker had failed to do so. He had the ability to link the most disparate observations into a coherent and testable working hypothesis. He effortlessly passed this trait on to many of his students and postdocs, who chose their research topics with very little interference from him. He did, however, reserve the right to challenge research plans. While I was an assistant professor in his lab, he suggested I set up a cell-free system to study the initial step in the secretory pathway — a task much easier said than done. But after two years of trying, I succeeded and it certainly made a big difference to my career.

He took considerable interest in the papers that were published by his lab. Even when he was not listed as an author, he meticulously edited and corrected each paper with his immaculate handwriting at the edge or on the opposite empty page of the typewritten manuscripts. I treasure the corrections he made on all my manuscripts during that time.

Palade moved to Yale in 1973, where he stayed until he joined the University of California, San Diego (UCSD) in 1990. At both universities, he continued to make many crucial discoveries and, as at Rockefeller, built thriving departments of cell biology. At UCSD, he served as the first dean of scientific affairs until his retirement at the age of 87.

Many of Palade's students and their second-, third- and fourth-generation 'descendants' are still major contributors to the field of cell biology. Among the prizes he was awarded are the Lasker prize, the Gairdner award and the Louisa Gross Horwitz Prize. He was also a joint recipient of the 1974 Nobel Prize in Physiology or Medicine.

Palade was deeply interested in music, the fine arts and history. He was an eloquent speaker, and his lectures are legendary examples of his lucidity and passion for his subject. He worked productively until his late eighties, when Parkinson's disease forced him to reduce his activities. It must have been hard for him to cope with these physical constraints, although his intellectual curiosity and passion remained intact for much longer. He is survived by his wife, Marilyn Farquhar, two children from his first marriage and two grandchildren.

Günter Blobel
Günter Blobel is at the Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.
e-mail: blobel@mail.rockefeller.edu

A. CAMERANO/AP

ARTICLES

Accurate whole human genome sequencing using reversible terminator chemistry

A list of authors and their affiliations appears at the end of the paper

DNA sequence information underpins genetic research, enabling discoveries of important biological or medical benefit. Sequencing projects have traditionally used long (400–800 base pair) reads, but the existence of reference sequences for the human and many other genomes makes it possible to develop new, fast approaches to re-sequencing, whereby shorter reads are compared to a reference to identify intraspecies genetic variation. Here we report an approach that generates several billion bases of accurate nucleotide sequence per experiment at low cost. Single molecules of DNA are attached to a flat surface, amplified *in situ* and used as templates for synthetic sequencing with fluorescent reversible terminator deoxyribonucleotides. Images of the surface are analysed to generate high-quality sequence. We demonstrate application of this approach to human genome sequencing on flow-sorted X chromosomes and then scale the approach to determine the genome sequence of a male Yoruba from Ibadan, Nigeria. We build an accurate consensus sequence from >30× average depth of paired 35-base reads. We characterize four million single-nucleotide polymorphisms and four hundred thousand structural variants, many of which were previously unknown. Our approach is effective for accurate, rapid and economical whole-genome re-sequencing and many other biomedical applications.

DNA sequencing yields an unrivalled resource of genetic information. We can characterize individual genomes, transcriptional states and genetic variation in populations and disease. Until recently, the scope of sequencing projects was limited by the cost and throughput of Sanger sequencing. The raw data for the three billion base (3 gigabase (Gb)) human genome sequence, completed in 2004 (ref. 1), was generated over several years for ~$300 million using several hundred capillary sequencers. More recently an individual human genome sequence has been determined for ~$10 million by capillary sequencing[2]. Several new approaches at varying stages of development aim to increase sequencing throughput and reduce cost[3–6]. They increase parallelization markedly by imaging many DNA molecules simultaneously. One instrument run produces typically thousands or millions of sequences that are shorter than capillary reads. Another human genome sequence was recently determined using one of these approaches[7]. However, much bigger improvements are necessary to enable routine whole human genome sequencing in genetic research.

We describe a massively parallel synthetic sequencing approach that transforms our ability to use DNA and RNA sequence information in biological systems. We demonstrate utility by re-sequencing an individual human genome to high accuracy. Our approach delivers data at very high throughput and low cost, and enables extraction of genetic information of high biological value, including single-nucleotide polymorphisms (SNPs) and structural variants.

DNA sequencing using reversible terminators

We generated high-density single-molecule arrays of genomic DNA fragments attached to the surface of the reaction chamber (the flow cell) and used isothermal 'bridging' amplification to form DNA 'clusters' from each fragment. We made the DNA in each cluster single-stranded and added a universal primer for sequencing. For paired read sequencing, we then converted the templates to double-stranded DNA and removed the original strands, leaving the complementary

strand as template for the second sequencing reaction (Fig. 1a–c). To obtain paired reads separated by larger distances, we circularized DNA fragments of the required length (for example, 2 ± 0.2 kb) and obtained short junction fragments for paired end sequencing (Fig. 1d).

We sequenced DNA templates by repeated cycles of polymerase-directed single base extension. To ensure base-by-base nucleotide incorporation in a stepwise manner, we used a set of four reversible terminators, 3′-*O*-azidomethyl 2′-deoxynucleoside triphosphates (A, C, G and T), each labelled with a different removable fluorophore (Supplementary Fig. 1a)[8]. The use of 3′-modified nucleotides allowed the incorporation to be driven essentially to completion without risk of over-incorporation. It also enabled addition of all four nucleotides simultaneously rather than sequentially, minimizing risk of misincorporation. We engineered the active site of 9°N DNA polymerase to improve the efficiency of incorporation of these unnatural nucleotides[9]. After each cycle of incorporation, we determined the identity of the inserted base by laser-induced excitation of the fluorophores and imaging. We added tris(2-carboxyethyl)phosphine (TCEP) to remove the fluorescent dye and side arm from a linker attached to the base and simultaneously regenerate a 3′ hydroxyl group ready for the next cycle of nucleotide addition (Supplementary Fig. 1b). The Genome Analyzer (GA1) was designed to perform multiple cycles of sequencing chemistry and imaging to collect the sequence data automatically from each cluster on the surface of each lane of an eight-lane flow cell (Supplementary Fig. 2).

To determine the sequence from each cluster, we quantified the fluorescent signal from each cycle and applied a base-calling algorithm. We defined a quality (Q) value for each base call (scaled as by the phred algorithm[10]) that represents the likelihood of each call being correct (Supplementary Fig. 3). We used the Q-values in subsequent analyses to weight the contribution of each base to sequence alignment and detection of sequence variants (for example, SNP

calling). We discarded all reads from mixed clusters and used the remaining 'purity filtered' reads for analysis. Typically we generated 1–2 Gb of high-quality purity filtered sequence per flow cell from ~30–60-million single 35-base reads, or 2–4 Gb in a paired read experiment (Supplementary Table 1).

To demonstrate accurate sequencing of human DNA, we sequenced a human bacterial artificial chromosome (BAC) clone (bCX98J21) that contained 162,752 bp of the major histocompatibility complex on
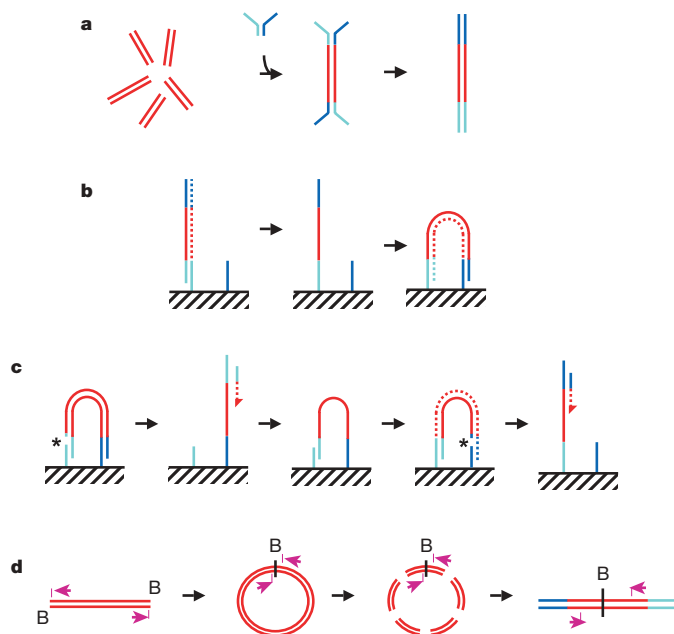

Figure 1 | Preparation of samples. a, DNA fragments are generated, for example, by random shearing and joined to a pair of oligonucleotides in a forked adaptor configuration. The ligated products are amplified using two oligonucleotide primers, resulting in double-stranded blunt-ended material with a different adaptor sequence on either end. **b**, Formation of clonal single-molecule array. DNA fragments prepared as in **a** are denatured and single strands are annealed to complementary oligonucleotides on the flow-cell surface (hatched). A new strand (dotted) is copied from the original strand in an extension reaction that is primed from the 3′ end of the surface-bound oligonucleotide; the original strand is then removed by denaturation. The adaptor sequence at the 3′ end of each copied strand is annealed to a new surface-bound complementary oligonucleotide, forming a bridge and generating a new site for synthesis of a second strand (dotted). Multiple cycles of annealing, extension and denaturation in isothermal conditions result in growth of clusters, each ~1 μm in physical diameter. This follows the basic method outlined in ref. 33. **c**, The DNA in each cluster is linearized by cleavage within one adaptor sequence (gap marked by an asterisk) and denatured, generating single-stranded template for sequencing by synthesis to obtain a sequence read (read 1; the sequencing product is dotted). To perform paired-read sequencing, the products of read 1 are removed by denaturation, the template is used to generate a bridge, the second strand is re-synthesized (shown dotted), and the opposite strand is then cleaved (gap marked by an asterisk) to provide the template for the second read (read 2). **d**, Long-range paired-end sample preparation. To sequence the ends of a long (for example, >1 kb) DNA fragment, the ends of each fragment are tagged by incorporation of biotinylated (B) nucleotide and then circularized, forming a junction between the two ends. Circularized DNA is randomly fragmented and the biotinylated junction fragments are recovered and used as starting material in the standard sample preparation procedure illustrated in **a**. The orientation of the sequence reads relative to the DNA fragment is shown (magenta arrows). When aligned to the reference sequence, these reads are oriented with their 5′ ends towards each other (in contrast to the short insert paired reads produced as shown in **a–c**). See Supplementary Fig. 17a for examples of both. Turquoise and blue lines represent oligonucleotides and red lines represent genomic DNA. All surface-bound oligonucleotides are attached to the flow cell by their 5′ ends. Dotted lines indicate newly synthesized strands during cluster formation or sequencing. (See Supplementary Methods for details.)

human chromosome 6 (accession AL662825.4, previously determined using capillary sequencing by the Wellcome Trust Sanger Institute). We developed a fast global alignment algorithm ELAND that aligns a read to the reference only if the read can be assigned a unique position with 0, 1 or 2 differences. We collected 0.17 Gb of aligned data for the BAC from one lane of a flow cell. Approximately 90% of the 35-base reads matched perfectly to the reference, demonstrating high raw read accuracy (Supplementary Fig. 4). To examine consensus coverage and accuracy, we used 5 Mb of 35-base purity filtered reads (30-fold average input depth of the BAC) and obtained 99.96% coverage of the reference. There was one consensus miscall, at a position of very low coverage (just above our cutoff threshold), yielding an overall consensus accuracy of >99.999%.

Detecting genetic variation of the human X chromosome

For an initial study of genetic variation, we sequenced flow-sorted X chromosomes of a Caucasian female (sample NA07340 originating from the Centre d'Etude du Polymorphisme Humain (CEPH)). We generated 278-million paired 30–35-bp purity filtered reads and aligned them to the human genome reference sequence. We carried out separate analyses of the data using two alignment algorithms: ELAND (see above) or MAQ (Mapping and Assembly with Qualities)[11]. Both algorithms place each read pair where it best matches the reference and assign a confidence score to the alignment. In cases where a read has two or more equally likely positions (that is, in an exact repeat), MAQ randomly assigns the read pair to one position and assigns a zero alignment quality score (these reads are excluded from SNP analysis). ELAND rejects all non-unique alignments, which are mostly in recently inserted retrotransposons (see Supplementary Fig. 5). MAQ therefore provides an opportunity to assess the properties of a data set aligned to the entire reference, whereas ELAND effectively excludes ambiguities from the short read alignment before further analysis.

We obtained comprehensive coverage of the X chromosome from both analyses. With MAQ, 204 million reads aligned to 99.94% of the X chromosome at an average depth of 43×. With ELAND, 192 million reads covered 91% of the reference sequence, showing what can be covered by unique best alignments. These results were obtained after excluding reads aligning to non-X sequence (impurities of flow sorting) and apparently duplicated read pairs (Supplementary Table 2). We reasoned that these duplicates (~10% of the total) arose during initial sample amplification.

The sampling of sequence fragments from the X chromosome is close to random. This is evident from the distribution of mapped read depth in the MAQ alignment in regions where the reference is unique (Fig. 2a): the variance of this distribution is only 2.26 times that of a Poisson distribution (the theoretical minimum). Half of this excess variance can be accounted for by a dependence on G+C content. However, the average mapped read depth only falls below 10× in regions with G+C content less than 4% or greater than 76%, comprising in total just 1% of unique chromosome sequence and 3% of coding sequence (Fig. 2b).

We identified 92,485 candidate SNPs in the X chromosome using ELAND (Supplementary Fig. 6). Most calls (85%) match previous entries in the public database dbSNP. Heterozygosity (π) in this data set is 4.3×10^{-4} (that is, one substitution per 2.3 kb), close to a previously published X chromosome estimate (4.7×10^{-4})[12]. Using MAQ we obtained 104,567 SNPs, most of which were common to the results of the ELAND analysis. The differences between the two sets of SNP calls are largely the consequence of different properties of the alignments as described earlier. For example, most of the SNPs found only by the MAQ-based analysis were at positions of low or zero sequence depth in the ELAND alignment (Supplementary Fig. 6c).

We assessed accuracy and completeness of SNP calling by comparison to genotypes obtained for this individual using the Illumina HumanHap550 BeadChip (HM550). The sequence data covered >99.8% of the 13,604 genotyped positions and we found excellent
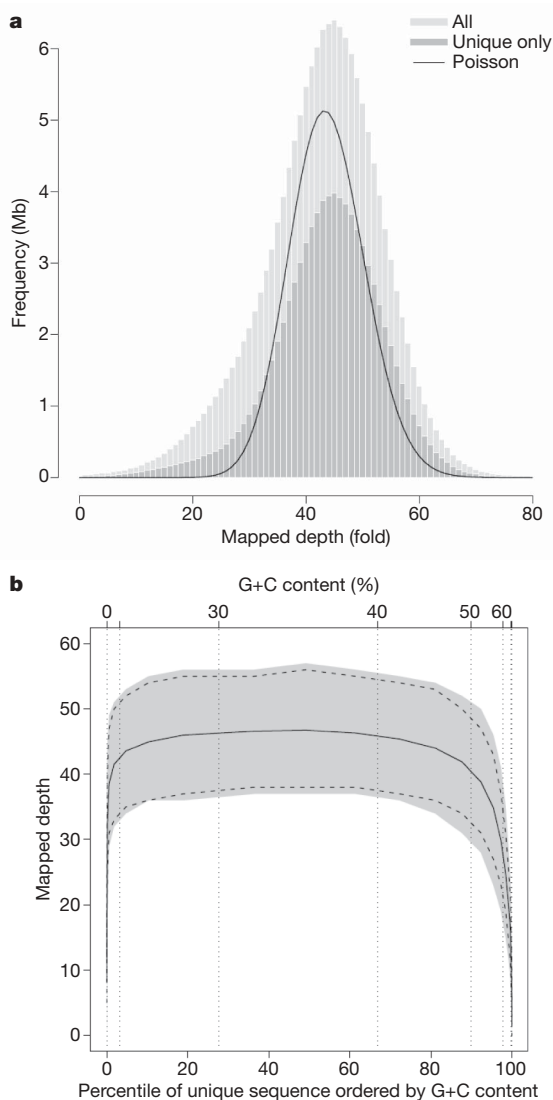
Figure 2 | X chromosome data. a, Distribution of mapped read depth in the X chromosome data set (NA07340), sampled at every 50th position along the chromosome and displayed as a histogram ('All'). An equivalent analysis of mapped read depth for the unique subset of these positions is also shown ('Unique only'). The solid line represents a Poisson distribution with the same mean. **b**, Distribution of X chromosome uniquely mapped reads as a function of G+C content. Note that the x axis is per cent G+C content and is scaled by percentile of unique sequence. The solid line is average mapped depth of unique sequence; the grey region is the central 80% of the data (10th to 90th centiles); the dashed lines are 10th and 90th centiles of a Poisson distribution with the same mean as the data.

agreement between sequence-based SNP calls and genotyping data (99.52% or 99.99% using ELAND or MAQ, respectively; Supplementary Table 3). There was complete concordance of all homozygous calls and a low level of 'under-calling' from the sequence data (denoted as 'GT>Seq' in Table 1) at a small number of the heterozygous sites, caused by inadequate sampling of one of the two alleles. The depth of input sequence influences the coverage and accuracy of SNP calling. We found that reducing the read depth to 15× still gives 97% coverage of genotype positions and only 1.27% of the heterozygous sites are under-called. We observed no other types of disagreement at any input depth (Supplementary Fig. 7).

We detected structural variants (defined as any variant other than a single base substitution) as follows. We found 9,747 short insertions/deletions ('short indels'; defined here as less than the length of the read) by performing a gapped alignment of individual reads (Supplementary Fig. 8). We identified larger indels based on read

depth and/or anomalous read pair spacing, similar to previous approaches[13–15]. We detected 115 indels in total, 77 of which were visible from anomalous read-pair spacing (see Supplementary Tables 4 and 5). We developed Resembl, an extension to the Ensembl browser[16], to view all variants (Supplementary Fig 9). Inversions can be detected when the orientation of one read in a pair is reversed (for example, see Supplementary Fig. 10). In general, inversions occur as the result of non-allelic homologous recombination, and are therefore flanked by repetitive sequence that can compromise alignments. We found partial evidence for other inversion events, but characterization of inversions from short read data is complex because of the repeats and requires further development.

Sequencing and analysis of a whole human genome

Our X chromosome study enabled us to develop an integrated set of methods for rapid sequencing and analysis of whole human genomes. We sequenced the genome of a male Yoruba from Ibadan, Nigeria (YRI, sample NA18507). This sample was originally collected for the HapMap project[17,18] through a process of community engagement and informed consent[19] and has also been studied in other projects[20,21]. We were therefore able to compare our results with publicly available data from the same sample. We constructed two libraries: one of short inserts (~200 bp) with similar properties to the previous X chromosome library and one from long fragments (~2 kb) to provide longer-range read-pair information (see Supplementary Fig. 11 for size distributions). We generated 135 Gb of sequence (~4 billion paired 35-base reads; see Supplementary Table 6) over a period of 8 weeks (December 2007 to January 2008) on six GA1 instruments averaging 3.3 Gb per production run (see Supplementary Table 1 for example). The approximate consumables cost (based on full list price of reagents) was $250,000. We aligned 97% of the reads using MAQ and found that 99.9% of the human reference (NCBI build 36.1) was covered with one or more reads at an average of 40.6-fold depth. Using ELAND, we aligned 91% of the reads over 93% of the reference sequence at sufficient depth to call a strong consensus (>three Q30 bases). The distribution of mapped read depth was close to random, with slight over-dispersion as seen for the X chromosome data. We observed comprehensive representation across a wide range of G+C content, dropping only at the very extreme ends, but with a different pattern of distribution compared to the X chromosome (see Supplementary Fig. 12).

We identified ~4 million SNPs, with 74% matching previous entries in dbSNP (Fig. 3). We found excellent agreement of our SNP calls with genotyping results: sequence-based SNP calls covered almost all of the 552,710 loci of HM550, with >99.5% concordance of sequencing versus genotyping calls (Table 1 and Supplementary Table 7a). The few disagreements were mostly under-calls of heterozygous positions (GT>Seq) in areas of low sequence depth, providing us with a false-negative rate of <0.35% from the ELAND analysis (see Table 1). The other disagreements (0.09% of all genotypes) included errors in genotyping plus apparent tri-allelic SNPs (Supplementary Table 7a). The main cause of genotype error (0.05% of all genotypes) is the existence of a second 'hidden' SNP close to the assayed locus that disrupts the genotyping assay, leading to loss of one allele and an erroneous homozygous genotype (Supplementary Figs 13 and 14).

To examine the accuracy of SNP calling in more detail, we compared our sequence-based SNP calls with 3.7 million genotypes (HM-All) generated for this sample during the HapMap project (Table 1 and Supplementary Table 7b)[18] and found excellent concordance between the data sets. Disagreements included sequence-based under-calls of heterozygous positions in regions of low read depth. The slightly higher level of other disagreements (0.76%) seen in this analysis compared to that of the HM550 data (0.09%) is in line with the higher level of underlying genotype error rate of 0.7% for the HapMap data[18]. To refine this analysis further, we generated a set of 530,750 very high confidence reference genotypes comprising

Table 1 | Comparison of SNP calls made from sequence versus genotype data for the human genome (NA18507) and X chromosome (NA07340)

	ELAND			MAQ				
	X	Human	Human	X	Human	Human	Human	
	HM550 (13,604 SNPs) (%)	HM550 (552,710 SNPs) (%)	HM-All (3,699,592 SNPs) (%)	HM550 (13,604 SNPs) (%)	HM550 (552,710 SNPs) (%)	HM-All (3,699,592 SNPs) (%)	Combined (530,750 SNPs)	
							(%)	(n)
Covered by sequence	99.77	99.60	99.24	99.91	99.74	99.29	99.78	529,589
Concordant calls	99.52	99.57	98.80	99.99	99.90	99.12	99.94	529,285
All disagreements	0.48	0.43	1.20	0.01	0.1	0.88	0.06	304
GT>Seq	0.48	0.35	0.46	0.01	0.03	0.15	0.02	130
Seq>GT	0	0.05	0.52	0	0.05	0.54	0.02	130
Other discordances	0	0.03	0.22	0	0.02	0.2	0.01	44

SNP panels referred to are HM550 (Illumina Infinium HumanHap550 BeadChip) and HM-All (complete data from phase 1 and phase 2 of the International HapMap Project). 'Combined' is a set of concordant genotypes from both sets (HM550 and HM-All; see text). GT>Seq denotes a heterozygous genotyping SNP call where there is a homozygous sequencing SNP call (one of the two alleles); Seq>GT denotes the converse (that is, a heterozygous sequencing SNP call where there is a homozygous genotyping call). Other discordances are differences in the two SNP calls that cannot be accounted for by one allele being missing from one call.

concordant calls in both the HM550 and HM-All genotype data sets. Comparing the results of the MAQ analysis to this high confidence set (see Table 1), we found 130 heterozygote under-calls GT>Seq (that is, a false-negative rate of 0.025%). There were also 130 heterozygote over-calls Seq>GT, but most of these are probably genotype errors as 82 have a nearby 'hidden' SNP and 3 have a nearby indel. A further 41 are tri-allelic loci, leaving at most 4 potential wrong calls by sequencing (that is, false-positive rate of 4 per 529,589 positions). Finally we selected a subset of novel SNP calls from the sequence data and tested them by genotyping. We found 96.1% agreement between sequence and genotype calls (Supplementary Table 8). However, the 47 disagreements included 10 correct sequencing calls (genotyping under-calls owing to hidden SNPs) and 7 sequencing under-calls. On this basis, therefore, the false-positive discovery rate for the one million novel SNPs is 2.5% (30 out of 1,206). For the entire data set of four million SNPs detected in this analysis, the false-positive and -negative rates both average <1%.

This genome from a Yoruba individual contains significantly more polymorphism than a genome of European descent. The autosomal heterozygosity (π) of NA18507 is 9.94×10^{-4} (1 SNP per 1,006 bp), higher than previous values for Caucasians (7.6×10^{-4}, ref. 12). Heterozygosity in the pseudoautosomal region 1 (PAR1) is substantially higher (1.92×10^{-3}) than the autosomal value. PAR1 (2.7 Mb)

at the tip of the short arm of chromosomes X and Y undergoes obligatory recombination in male meiosis, which is equivalent to $20\times$ the autosome average. This illustrates a clear correlation between recombination and nucleotide diversity. By contrast, the 0.33-Mb PAR2 region has a much lower recombination rate than PAR1; we observed that heterozygosity in PAR2 is identical to that of the autosomes in NA18507. Heterozygosity in coding regions is lower (0.54×10^{-3}) than the total autosome average, consistent with the model that some coding changes are deleterious and are lost as the result of natural selection[22]. Nevertheless, the 26,140 coding SNPs (Supplementary Fig. 15) include 5,361 non-conservative amino acid substitutions plus 153 premature termination codons (Supplementary Table 9), many of which are expected to affect protein function.

We performed a genome-wide survey of structural variation in this individual and found excellent correlation with variants that had been reported in previous studies, as well as detecting many new variants. We found 0.4 million short indels (1–16 bp; Supplementary Fig. 16), most of which are length polymorphisms in homopolymeric tracts of A or T. Half of these events are corroborated by entries in dbSNP, and 95 of 100 examined were present in amplicons sequenced from this individual in ENCODE regions, confirming the high specificity of this method of short indel detection. For larger structural variants (detected by anomalously spaced paired ends) we found that some were detected by both long and short insert data sets (Supplementary Fig. 17a), but most were unique to one or other data set. We observed two reasons for this: first, small events (<400 bp) are within the normal size variance of the long insert data; second, nearby repetitive structures can prevent unique alignment of read pairs (see Supplementary Fig. 17b, c). In some cases, the high resolution of the short insert data permits detection of additional complexity in a structural rearrangement that is not revealed by the long insert data. For example, where the long insert data indicate a 1.3-kb deletion in NA18507 relative to the reference, the short insert data reveal an inversion accompanied by deletions at both breakpoints (Fig. 4). We carried out *de novo* assembly of reads in this region and constructed a single contig that defines the exact structure of the rearrangement (data not shown).

We discovered 5,704 structural variants ranging from 50 bp to >35 kb where there is sequence absent from the genome of NA18507 compared to the reference genome. We observed a steadily decreasing number of events of this type with increasing size, except for two peaks (Supplementary Fig. 18). Most of the events represented by the large peak at 300–350 bp contain a sequence of the AluY family. This is consistent with insertion of short interspersed nuclear elements (SINEs) that are present in the reference genome but missing from the genome of NA18507. Similarly, the second, smaller peak at 6–7 kb is the consequence of insertion of the long interspersed nuclear element (LINE) L1 *Homo sapiens* (L1Hs) in
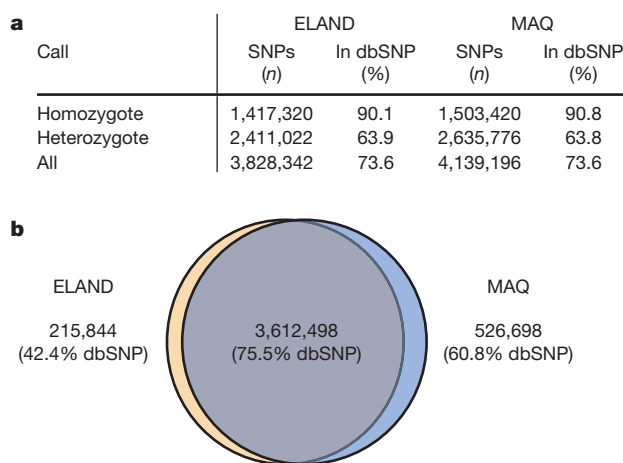
a		ELAND		MAQ	
Call		SNPs (n)	In dbSNP (%)	SNPs (n)	In dbSNP (%)
Homozygote		1,417,320	90.1	1,503,420	90.8
Heterozygote		2,411,022	63.9	2,635,776	63.8
All		3,828,342	73.6	4,139,196	73.6

b

ELAND
215,844
(42.4% dbSNP)

3,612,498
(75.5% dbSNP)

MAQ
526,698
(60.8% dbSNP)

Figure 3 | SNPs identified in the human genome sequence of NA18507.
a, Number of SNPs detected by class and percentage in dbSNP (release 128). Results from ELAND and MAQ alignments are reported separately.
b, Analysis of SNPs detected in each analysis reveals extensive overlap. The percentage of NA18507 SNP calls that match previous entries in dbSNP is lower than that of our X chromosome study (see Supplementary Fig. 6). We expect this because individual NA07340 (from the X chromosome study) was also previously used for discovery and submission of SNPs to dbSNP during the HapMap project, in contrast to NA18507.
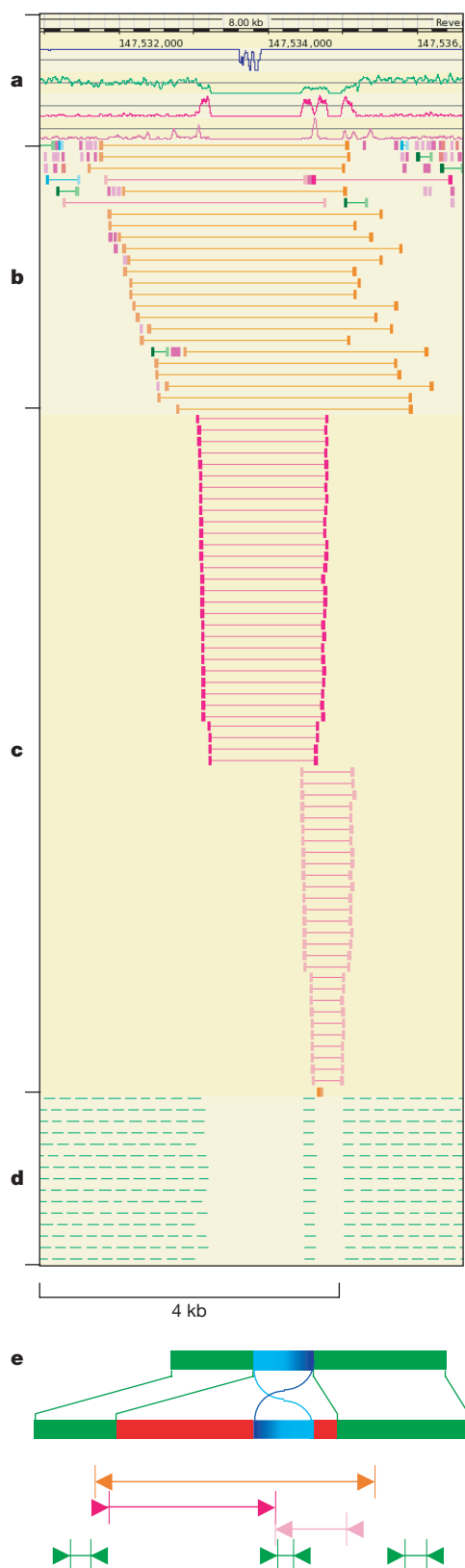
Figure 4 | Homozygous complex rearrangement detected by anomalous paired reads. The rearrangement involves an inversion of 369 bp (blue–turquoise bar in the schematic diagram) flanked by deletions (red bars) of 1,206 and 164 bp, respectively, at the left- and right-hand breakpoints. **a**, Summary tracks in the Resembl browser, denoting scale, simulated alignability of reads to reference (blue plot), actual aligned depth of coverage by NA18507 reads (green plot), density of anomalous reads indicating structural variants (red plot; peaks denote 'hotspots') and density of singleton reads (pink plot). **b**, Anomalous long-insert read pairs (orange lines denote DNA fragment; blocks at either end denote each read); the data indicate loss of ~1.3 kb in NA18507 relative to the reference. **c**, Anomalous short-insert pairs of two types (red and pink) indicate an inverted sequence flanked by two deletions. **d**, Normal short-insert read-pair alignments (each green line denotes the extent of the reference that is covered by the short fragment, including the two reads). **e**, The schematic diagram depicts the arrangement of normal and anomalous read pairs relative to the rearrangement. Top line, structure of NA18507; second line, structure of reference sequence. Green bars denote sequence that is collinear in the reference and NA18507 genomes. The turquoise–blue bar illustrates the inverted segment. Red bars indicate the sequences present in the reference but absent in NA18507. Arrows denote orientation of reads when aligned to the reference. The display in **a**–**d** is a composite of screen shots of the same window, overlapped for display purposes.

Supplementary Fig. 20. The 'singleton' reads on either side of the event, which have partners that do not align to the reference, form part of a *de novo* assembly that precisely defines the novel sequence and breakpoint (Supplementary Fig. 21).

Effect of sequence depth on coverage and accuracy

We investigated the impact of varying input read depth (and hence cost) on SNP calling using chromosome 2 as a model. SNP discovery increases with increasing depth: essentially all homozygous positions are detected at 15×, whereas heterozygous positions accumulate more gradually to 33× (Fig. 5a). This effect is influenced by the stringency of the SNP caller. To call each allele in this analysis we required the equivalent of two high-quality Q30 bases (as opposed to three used in full depth analyses). Homozygotes could be detected at read depth of 2× or higher, whereas heterozygote detection required at least double this depth for sampling of both alleles. Missing calls (not covered by sequence) and discordances between sequence-based SNP calls and genotype loci (mostly under-calls of heterozygotes due to low depth) progressively reduced with increasing depth (Fig. 5b). We observed very few other types of discordance at any depth; many of these are genotyping errors as described above.

Concluding remarks

Reversible terminator chemistry is a defining feature of this sequencing approach, enabling each cycle to be driven to completion while minimizing misincorporation. The result is a system that generates accurate data at very high throughput and low cost. We determined an accurate whole human genome sequence in 8 weeks to an average depth of ~40×. We built a consensus sequence, optimized methods for analysis, assessed accuracy and characterized the genetic variation of this individual in detail.

We assessed accuracy relative to genotype data over the entire fraction of the human sequence where SNP calling was possible (>90%). We established very low false-positive and -negative rates for the ~four million SNPs detected (<1% over-calls and under-calls). This compares favourably with previous individual genome analyses which reported a 24% under-calling of heterozygous positions[2,7].

Paired reads were very powerful in all areas of the analysis. They provided very accurate read alignment and thus improved the accuracy and coverage of consensus sequence and SNP calling. They were essential for developing our short indel caller, and for detecting larger structural variants. Our short-insert paired-read data set introduced a new level of resolution in structural variation detection, revealing thousands of variants in a size range not characterized previously. In

many cases. We found good correspondence between our results and the data of ref. 23, which reported 148 deletions of <100 kb in this individual on the basis of abnormal fosmid paired-end spacing. We found supporting evidence for 111 of these events. We detected a further 2,345 indels in the range 60–160 bp which are sequences present in the genome of NA18507 and absent from the reference genome (Supplementary Fig. 19). One example is shown in

a SNP calls versus sequence depth

b Missing or discordant data versus sequence depth
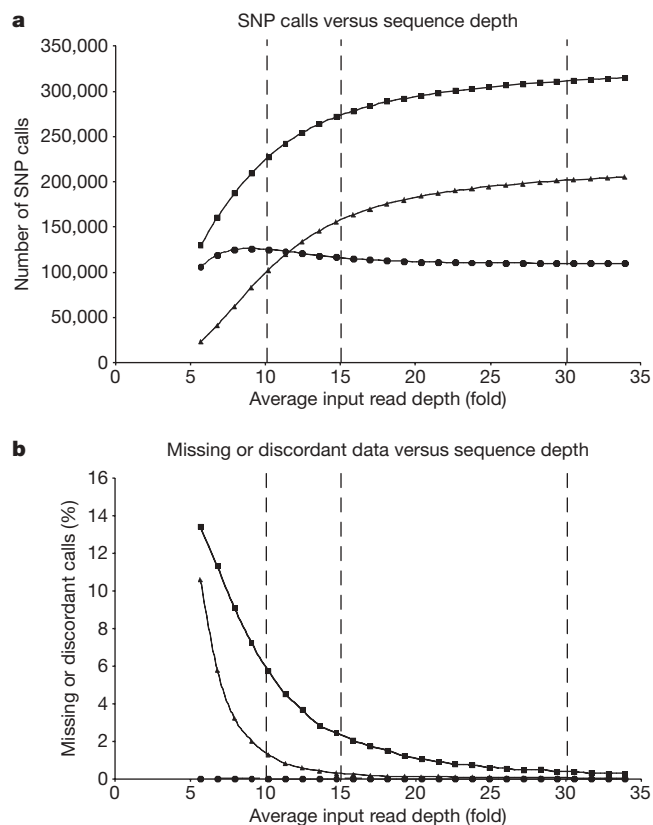
Figure 5 | Effect of sequence depth on coverage and accuracy of human genome sequencing. ELAND alignments were used for this analysis. **a**, Accumulation of sequence-based SNP calls, including all SNPs (squares), heterozygous SNPs (triangles) and homozygous SNPs (circles) with increasing input read depth. **b**, Decrease in genotype positions not covered by sequence (squares), heterozygote under-calls in sequence data relative to genotype data (triangles) and discordant SNP calls compared to genotypes (circles) with increasing input read depth. Vertical dotted lines indicate various input read depths ($10\times$, $15\times$, $30\times$ haploid genome).

filtered read data are available for download from the Short Read Archive at NCBI or from the European Short Read Archive (ERA) at the EBI.

Analysis software. Image analysis software and the ELAND aligner are provided as part of the Genome Analyzer analysis software. SNP and structural variant detectors will be available as future upgrades of the analysis pipeline. The Resembl extension to Ensembl is available on request. The MAQ (Mapping and Assembly with Qualities) aligner is freely available for download from http://maq.sourceforge.net.

Data access. Sequence data for NA18507 are freely available from the NCBI short read archive, accession SRA000271 (ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/SRA000271). X chromosome data are freely available from ERA, accession ERA000035. Links to Resembl displays for chromosome X and human data, plus information on other available data, are provided at http://www.illumina.com/HumanGenome.

See Supplementary Methods for a detailed Methods section.

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
3. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
4. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
5. Harris, T. D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
6. Lundquist, P. M. *et al.* Parallel confocal detection of single molecules in real time. *Opt. Lett.* **33**, 1026–1028 (2008).
7. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
8. Milton, J. *et al.* Modified nucleotides. World Intellectual Property Organization WO/2004/018497 (2004).
9. Smith, G. P. *et al.* Modified polymerases for improved incorporation of nucleotide analogues. World Intellectual Property Organization WO/2005/024010 (2005).
10. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
11. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* doi:10.1101/gr.078212.108 (25 September 2008).
12. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
13. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
14. Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
15. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
16. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).
17. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
18. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
19. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
20. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
21. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
22. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
23. Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
24. Hillier, L. W. *et al.* Whole-genome sequencing and variant discovery in C. elegans. *Nature Methods* **5**, 183–188 (2008).
25. Hodges, E. *et al.* Genome-wide in situ exon capture for selective resequencing. *Nature Genet.* **39**, 1522–1527 (2007).
26. Porreca, G. J. *et al.* Multiplex amplification of large sets of human exons. *Nature Methods* **4**, 931–936 (2007).
27. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
28. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
29. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

some cases we determined the exact sequence of structural variants by *de novo* assembly from the same paired-read data set. Interpreting events that are embedded in repetitive sequence tracts will require further work.

Massively parallel sequencing technology makes it feasible to consider whole human genome sequencing as a clinical tool in the near future. Characterizing multiple individual genomes will enable us to unravel the complexities of human variation in cancer and other diseases and will pave the way for the use of personal genome sequences in medicine and healthcare. Accuracy of personal genetic information from sequence will be critical for life-changing decisions.

In addition to the large-scale genomic projects exemplified by the present study and others[15,24–26], the system described here is being used to explore biological phenomena in unprecedented detail, including transcriptional activity, mechanisms of gene regulation and epigenetic modification of DNA and chromatin[27–32]. In the future, DNA sequencing will be the central tool for unravelling how genetic information is used in living processes.

METHODS SUMMARY

DNA and sequencing. DNA samples (NA07340 and NA18507) and cell line (GM07340) were obtained from Coriell Repositories. DNA samples were genotyped on the HM550 array and the results compared to publicly available data to confirm their identity before use. Methods for DNA manipulation, including sample preparation, formation of single-molecule arrays, cluster growth and sequencing were all developed during this study and formed the basis for the standard protocols now available from Illumina, Inc. All sequencing was performed on Illumina GA1s equipped with a one-megapixel camera. All purity

30. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).

31. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).

32. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 585–587 (2008).

33. Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* **34**, e22 (2006).

David R. Bentley[1], Shankar Balasubramanian[2], Harold P. Swerdlow[1]†, Geoffrey P. Smith[1], John Milton[1]†, Clive G. Brown[1]†, Kevin P. Hall[1], Dirk J. Evers[1], Colin L. Barnes[1,2], Helen R. Bignell[1], Jonathan M. Boutell[1], Jason Bryant[1], Richard J. Carter[1], R. Keira Cheetham[1], Anthony J. Cox[1], Darren J. Ellis[1], Michael R. Flatbush[3], Niall A. Gormley[1], Sean J. Humphray[1], Leslie J. Irving[1], Mirian S. Karbelashvili[3], Scott M. Kirk[3], Heng Li[4], Xiaohai Liu[1,2], Klaus S. Maisinger[1], Lisa J. Murray[1], Bojan Obradovic[1], Tobias Ost[1], Michael L. Parkinson[1], Mark R. Pratt[3], Isabelle M. J. Rasolonjatovo[1], Mark T. Reed[3], Roberto Rigatti[1], Chiara Rodighiero[1], Mark T. Ross[1], Andrea Sabot[1], Subramanian V. Sankar[3], Aylwyn Scally[4], Gary P. Schroth[3], Mark E. Smith[1], Vincent P. Smith[1], Anastassia Spiridou[1], Peta E. Torrance[1], Svilen S. Tzonev[1], Eric H. Vermaas[3], Klaudia Walter[4], Xiaolin Wu[1], Lu Zhang[3], Mohammed D. Alam[3], Carole Anastasi[1], Ify C. Aniebo[1], David M. D. Bailey[1], Iain R. Bancarz[1], Saibal Banerjee[3], Selena G. Barbour[1], Primo A. Baybayan[3], Vincent A. Benoit[1], Kevin F. Benson[1], Claire Bevis[1], Phillip J. Black[1], Asha Boodhun[1], Joe S. Brennan[1], John A. Bridgham[3], Rob C. Brown[1], Andrew A. Brown[1], Dale H. Buermann[3], Abass A. Bundu[1], James C. Burrows[3], Nigel P. Carter[4], Nestor Castillo[3], Maria Chiara E. Catenazzi[1], Simon Chang[3], R. Neil Cooley[1], Natasha R. Crake[1], Olubunmi O. Dada[1], Konstantinos D. Diakoumakos[1], Belen Dominguez-Fernandez[1], David J. Earnshaw[1,2], Ugonna C. Egbujor[1], David W. Elmore[3], Sergey S. Etchin[3], Mark R. Ewan[3], Milan Fedurco[5], Louise J. Fraser[1], Karin V. Fuentes Fajardo[1], W. Scott Furey[2], David George[3], Kimberley J. Gietzen[6], Colin P. Goddard[1], George S. Golda[3], Philip A. Granieri[3], David E. Green[1], David L. Gustafson[3], Nancy F. Hansen[7], Kevin Harnish[1], Christian D. Haudenschild[1], Narinder I. Heyer[1], Matthew M. Hims[1], Johnny T. Ho[3], Adrian M. Horgan[1], Katya Hoschler[1], Steve Hurwitz[3], Denis V. Ivanov[3], Maria Q. Johnson[3], Terena James[1], T. A. Huw Jones[1], Gyoung-Dong Kang[1], Tzvetana H. Kerelska[3], Alan D. Kersey[1], Irina Khrebtukova[3], Alex P. Kindwall[3], Zoya Kingsbury[1], Paula I. Kokko-Gonzales[1], Anil Kumar[1], Marc A. Laurent[6], Cynthia T. Lawley[6], Sarah E. Lee[1], Xavier Lee[1], Arnold K. Liao[3], Jennifer A. Loch[1], Mitch Lok[3], Shujun Luo[3], Radhika M. Mammen[1], John W. Martin[3], Patrick G. McCauley[1], Paul McNitt[3], Parul Mehta[1], Keith W. Moon[3], Joe W. Mullens[3], Taksina Newington[1], Zemin Ning[4], Bee Ling Ng[4], Sonia M. Novo[1], Michael J. O'Neill[3], Mark A. Osborne[1,2], Andrew Osnowski[1], Omead Ostadan[3,6], Lambros L. Paraschos[3], Lea Pickering[1], Andrew C. Pike[1], Alger C. Pike[3], D. Chris Pinkard[3], Daniel P. Pliskin[3], Joe Podhasky[3], Victor J. Quijano[3], Come Raczy[1], Vicki H. Rae[1], Stephen R. Rawlings[1], Ana Chiva Rodriguez[1], Phyllida M. Roe[1], John Rogers[1], Maria C. Rogert Bacigalupo[1], Nikolai Romanov[1], Anthony Romieu[5], Rithy K. Roth[3], Natalie J. Rourke[1], Silke T. Ruediger[3], Eli Rusman[3], Raquel M. Sanches-Kuiper[1], Martin R. Schenker[1], Josefina M. Seoane[3], Richard J. Shaw[1], Mitch K. Shiver[3], Steven W. Short[3], Ning L. Sizto[3], Johannes P. Sluis[3], Melanie A. Smith[1], Jean Ernest Sohna Sohna[1], Eric J. Spence[3], Kim Stevens[1], Neil Sutton[1], Lukasz Szajkowski[1], Carolyn L. Tregidgo[1], Gerardo Turcatti[5], Stephanie vandeVondele[1], Yuli Verhovsky[3], Selene M. Virk[3], Suzanne Wakelin[3], Gregory C. Walcott[3], Jingwen Wang[3], Graham J. Worsley[1], Juying Yan[3], Ling Yau[3], Mike Zuerlein[3], Jane Rogers[4]†, James C. Mullikin[7], Matthew E. Hurles[4], Nick J. McCooke[1]†, John S. West[3], Frank L. Oaks[3], Peter L. Lundberg[3], David Klenerman[2], Richard Durbin[4] & Anthony J. Smith[1]

[1]Illumina Cambridge Ltd. (Formerly Solexa Ltd), Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK. [2]Department of Chemistry, University of Cambridge, The University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK. [3]Illumina Hayward (Formerly Solexa Inc.), 23851 Industrial Boulevard, Hayward, California 94343, USA. [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [5]Manteia Predictive Medicine S.A. Zone Industrielle, Coinsins, CH-1267, Switzerland. [6]Illumina Inc., Corporate Headquarters, 9883 Towne Centre Drive, San Diego, California 92121, USA. [7]National Human Genome Research Institute, National Institutes of Health, 41 Center Drive, MSC 2132, 9000 Rockville Pike, Bethesda, Maryland 20892-2132, USA. †Present addresses: The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK (H.P.S.); Oxford Nanopore Technologies, Begbroke Science Park, Sandy Lane, Kidlington OX5 1PF, UK (J.M., C.G.B.); BBSRC Genome Analysis Centre, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK (J.R.); Pronota, NV, VIB Bio-Incubator, Technologiepark 4, B-9052 Zwijnaarde/Ghent, Belgium (N.J.M.).

ARTICLES

The diploid genome sequence of an Asian individual

Jun Wang[1,2,3,4]*, Wei Wang[1,3]*, Ruiqiang Li[1,3,4]*, Yingrui Li[1,5,6]*, Geng Tian[1,7], Laurie Goodman[1], Wei Fan[1], Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Binxiao Feng[1], Heng Li[1,8], Yao Lu[1], Xiaodong Fang[1], Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Yujie Hu[1,7], Zhenzhen Yang[1], Hancheng Zheng[1], Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing Zhao[1], Jinjie Duan[1], Yan Zhou[1], Junjie Qin[1,7], Lijia Ma[1,7], Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Chang Yu[1], Fang Liang[1,7], Wenjie Li[1], Shaochuan Li[1], Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zhu[1], Dongyuan Liu[1], Zhike Lu[1], Ning Li[1,7], Guangwu Guo[1,7], Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,7], Quan Chen[1,5], Yu Liang[1,7], Yeyang Su[1,7], A. san[1,7], Cuo Ping[1,7], Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Zheng[1,4], Yuanyuan Ren[1], Ling Yang[1], Yang Gao[1,6], Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen[4], Gane Ka-Shu Wong[1,10], Rasmus Nielsen[9], Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songgang Li[1,2,5], Huanming Yang[1,2,3] & Jian Wang[1,2,3]

Here we present the first diploid genome sequence of an Asian individual. The genome was sequenced to 36-fold average coverage using massively parallel sequencing technology. We aligned the short reads onto the NCBI human reference genome to 99.97% coverage, and guided by the reference genome, we used uniquely mapped reads to assemble a high-quality consensus sequence for 92% of the Asian individual's genome. We identified approximately 3 million single-nucleotide polymorphisms (SNPs) inside this region, of which 13.6% were not in the dbSNP database. Genotyping analysis showed that SNP identification had high accuracy and consistency, indicating the high sequence quality of this assembly. We also carried out heterozygote phasing and haplotype prediction against HapMap CHB and JPT haplotypes (Chinese and Japanese, respectively), sequence comparison with the two available individual genomes (J. D. Watson and J. C. Venter), and structural variation identification. These variations were considered for their potential biological impact. Our sequence data and analyses demonstrate the potential usefulness of next-generation sequencing technologies for personal genomics.

The completion of a highly refined, encyclopaedic human genome sequence[1,2] was a major scientific development. Such reference sequences have accelerated human genetic analyses and contributed to advances in biomedical research. Given the growth of information on genetic risk factors, researchers are developing new tools and analyses for deciphering the genetic composition of a single person to refine medical intervention at a level tailored to the individual. The announcements that J. Craig Venter and James D. Watson have had their genomes sequenced[3,4], along with the announcement of the Personal Genome Project[5], highlight the growth of personal genomics.

Using a massively parallel DNA sequencing method, we have generated the first diploid genome sequence of a Han Chinese individual, a representative of an East Asian population that accounts for nearly 30% of the human population. The consensus sequence of the donor, assembled as pseudo-chromosomes, serves as one of the first sequences available from a non-European population and adds to the small number of publicly available individual genome sequences. This sequence and the analyses herein provide an initial step towards attaining information on population and individual genetic variation, and, given the use and analysis of next-generation sequencing

technology, constitute advancement towards the goal of providing personalized medicine.

Data production and short read alignment

The genomic DNA used in this study came from an anonymous male Han Chinese individual who has no known genetic diseases. The donor gave written consent for public release of the genomic data for use in scientific research (see Supplementary Information for consent forms).

We carried out G-banded karyotyping to check the overall structural suitability of this DNA for use as a genomic standard for other genetic comparison and found no obvious chromosomal abnormalities (Supplementary Fig. 1). We then proceeded with whole-genome sequencing of the individual's DNA (hereafter referred to as YH) using Illumina Genome Analysers (GA; see Methods for details). To minimize the likelihood of systematic biases in genome representation, multiple DNA libraries were prepared and data were generated from eight single-end and two paired-end libraries (Supplementary Table 1). The read lengths averaged 35 base pairs (bp), and the two paired-end libraries had a span size of 135 bp

and 440 bp, respectively. We collected a total of 3.3 billion reads of high-quality data: approximately 117.7 gigabases (Gb) of sequence (72 Gb from single-end reads and 45.7 Gb from paired-end reads). The data have been deposited in the EBI/NCBI Short Read Archive (accession number ERA000005). (See Supplementary Information for details concerning the availability of all data.)

Using the Short Oligonucleotide Alignment Program (SOAP)[6], 102.9 Gb of sequence (87.4% of all data) was properly aligned to the NCBI human reference genome (build 36.1; hereafter called NCBI36). This resulted in a 36-fold average coverage of NCBI36 (Table 1). The effective genome coverage of the single- and paired-end sequencing was 22.5-fold and 13.5-fold, respectively. In total, 99.97% of NCBI36 (excluding Ns, which are undetermined sequence of the reference genome) was covered by at least one uniquely or repeatedly aligned read (uniquely aligned reads had only one best hit on NBCI36; repeatedly aligned reads had multiple possible alignments; see Methods for details).

About 86.1% (83.6% of single-end and 90.2% of paired-end reads) of the mapped reads could be uniquely aligned and had an average per-nucleotide difference of 1.45% from the NCBI36 sequence. (See Supplementary Information for additional sequence alignment assessment.) We used the alignment of uniquely mapped single-end and paired-end reads to build the consensus YH genome sequence and to detect genetic variations: SNPs, insertions and deletions (indels), and structural variations.

SNP and indel identification

For SNP identification, we estimated the genotype and its accuracy for each nucleotide using Bayesian theory with probabilities based on previous observation of a SNP at that site. Each location was assigned a score value as a measure of SNP call accuracy (see Methods for details).

For SNP detection, we used a series of filtering criteria (see Methods) to remove unreliable portions of the consensus sequence from the analysis. The resulting calculated YH genome consensus sequence covered 92% of the NCBI36 sequence (92.6% of the autosomes; 83.1% of the sex chromosomes), in which we identified 3.07 million SNPs. The remaining 8% of the reference sequence was composed of either repetitive sequence (6.6%) that did not have any uniquely mapped reads or sequence that didn't pass our filtering steps (1.4%).

For indel identification, we required at least three pairs of reads to define an indel. We only considered paired-end read-gapped alignments that had insertion or deletion sizes of 3 bp or less to avoid creating alignment errors. Confining indel size was necessary to obtain the best detection accuracy given our short-read sequencing strategy. From this analysis, we identified a total of 135,262 indels.

SNP and indel identification accuracy

We assessed our SNP calling accuracy by comparing the identified SNPs in the YH sequence with dbSNP[7]. We found that 2.26 million (73.5%) of the YH SNPs were present in dbSNP as validated SNPs, and 0.4 million (12.9%) were present as non-validated SNPs. The remaining 0.42 million SNPs were novel (Fig. 1a). Of the 135,262 small indels that we identified, the percentage that overlapped dbSNP indels was much lower than that of the YH SNPs (40.9% compared
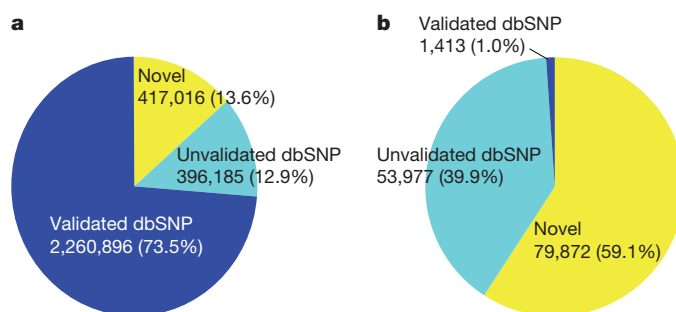
Figure 1 | The percentage of detected SNPs (a) **and small indels** (b) **that overlap with SNPs and small indels in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/, build 128).** The dbSNP alleles were separated into validated and non-validated SNPs, and the detected SNPs that were not present in dbSNP were classified as novel.

with 86.4%, respectively). Additionally, most (59.1%) of the indels were novel (Fig. 1b). This isn't surprising given that dbSNP contains only 13,727 validated and 1,589,264 non-validated 1–3-bp indels.

We also used the Illumina 1M BeadChip for genotyping. The YH consensus sequence covered 99.22% of the genotyped SNPs with a concordance rate at 99.90% (Table 2). We used polymerase chain reaction (PCR) amplification and traditional Sanger sequencing technology on a subset of the inconsistent SNPs and small indels to determine whether they conformed to the genotyping or GA sequencing results (Supplementary Table 2). Of the 50 SNPs examined, 82.0% (41 SNPs) were consistent with the GA sequencing, indicating that the YH genome has a 99.98% accuracy over these genotyped sites (Supplementary Table 3). We also validated 100% of the PCR-amplified YH genome non-coding-region indels and 90% of the frameshift indels (Supplementary Table 4).

Depth effect on genome sequencing

To determine what sequencing depth provides the best genome coverage and lowest SNP-calling error rates for a diploid human genome, we randomly extracted subsets of reads with different average depths from all the mapped reads on chromosome 12, which has a relatively moderate number of repeats. SNPs were identified using GA sequencing and then compared with the genotyping data. We applied the same filtering steps as used in SNP identification (see Methods).

At a depth greater than 10-fold, the assembled consensus covered 83.63% of the reference genome using single-end reads and 95.88% coverage using paired-end reads. Thus, greater sequencing depth provides only a small increase in genome coverage (Fig. 2).

The error rate of SNP calling, however, greatly decreases with increased sequencing depth. Additionally, the use of paired-end reads as opposed to single-end reads further reduces SNP calling errors. Of note, SNP calling errors of homozygous and heterozygous SNPs differ significantly.

Individual genome comparison

With the availability of the YH genome sequence, there are now three different individual genome sequences that can be compared. In looking at the SNPs of the three individual genomes, all share 1.2

Table 1 | Data production and alignment results for the YH genome

Data type	Number of reads	Number of mapped reads	Total bases (Gb)	Mapped bases (Gb)	Effective depth (fold)	Percentage with unique placement	Rate of nucleotide mismatches (%)
SE	2,019,025,890	1,921,271,902	72	64.4	22.5	83.60	1.62
PE	1,315,249,404	1,028,695,924	45.7	38.5	13.5	90.20	1.16
Total	3,334,275,294	2,949,967,826	117.7	102.9	36	86.10	1.45

Single-end (SE) and paired-end (PE) sequencing reads were aligned onto the reference assembly in NCBI build 36.1, allowing at most two mismatches or one continuous gap with a size of 1–3 bp. Effective depth was determined through the calculation of all mapped bases divided by the length of NCBI36 (excluding Ns, 2,858,013,089 bp in length). 'Unique placement' means a read had only one best placement with the least number of mismatches and gaps. The rate of nucleotide mismatches is the percentage of mismatched nucleotides over all mapped nucleotides, including sequencing errors and real genetic variations. In total, 487 million reads (14.6%) could not be aligned to the reference genome.

Table 2 | Comparison of GA sequencing and Illumina 1M genotyping alleles

	Allele type		Illumina 1M genotyping				Total	Consistency (%)
			HOM ref.	HOM mut.	HET ref.	HET mut.		
GA consensus	HOM ref.	2	566,825	–	–	–	567,266	99.92
		1	–	–	227	–		
		0	–	205	–	9		
	HOM mut.	2	–	217,179	–	–	217,242	99.97
		1	–	–	24	0		
		0	32	7	0	0		
	HET ref.	2	–	–	245,749	–	246,314	99.77
		1	289	252	24	0		
		0	–	0	–	0		
	HET mut.	2	–	–	–	0	22	0
		1	–	14	0	8		
		0	0	0	0	–		
	Missing		1,789	1,658	4,626	0	8,073	–
	Total		568,935	219,315	250,650	17	1,038,917	99.90
	Coverage (%)		99.69	99.24	98.15	100	99.22	–

We classified both the array-based genotyped alleles and the alleles that were called by the Illumina Genome Analyser (GA) into four categories: (1) HOM ref. (homozygotes where both alleles are identical to the reference); (2) HOM mut. (homozygotes where both alleles differ from the reference); (3) HET ref. (heterozygotes where only one allele is identical to the reference); and (4) HET mut. (heterozygotes where both alleles differ from the reference and also differ from one another). The number of GA sequencing sites that are consistent with genotyping at both alleles, at one allele, or that are inconsistent at both alleles were categorized as 2, 1, and 0, respectively. The genotyping array primarily included the major alleles of the most common SNPs found in the human population, so very few alleles found in the BeadChip analysis were sorted into category 4.

million SNPs. Each also has a set of SNPs unique to their own genome: for YH, 978,370 (31.8%) SNPs; for Venter, 924,333 (30.1%); and for Watson, 1,096,873 (33.0%) (Supplementary Fig. 2).

The three individuals also have a similar fraction of non-synonymous SNPs (YH, 7,062 (0.23%); Venter, 6,889 (0.22%); Watson, 7,319 (0.20%)). There are 2,622 non-synonymous SNPs shared among the three individuals, accounting for 37.1% of non-synonymous SNPs in the YH genome.

Mutation and selection

To determine which are the ancestral versions of the small indels between the YH genome and the NCBI reference genome, we used the chimpanzee genome as an outgroup and assumed that the alleles on the chimpanzee genome were the ancestral type (Supplementary Table 5). Notably, the YH genome has the ancestral version of 66.2% of the homozygous insertions, whereas the NCBI reference genome contained the ancestral versions of 66.0% of the homozygous deletions. This suggests that during the process of mutation and selection of the human genome, small DNA deletions occur more frequently than do small DNA insertions. Among the heterozygous indels, the allele types that are identical to those in the NCBI reference were mostly comprised of the ancestral versions. This is probably because alleles that are identical between two random individuals are more

Figure 2 | Genome coverage of the assembled consensus sequence and the accuracy of SNP detection as a function of sequencing depth. Analyses were carried out on human chromosome 12, and subsets of reads from all mapped 22.5× single-end and 13.5× paired-end reads were randomly extracted from areas of different average depth. The same method and filtering threshold (Q20) was used for SNP detection over different sequencing depths. The error rate for SNP calling—the sum of 'over call', 'under call' and 'misses' rate (see Supplementary Information)—was separated into heterozygotes (HET) and homozygotes (HOM), and was validated against the Illumina 1M genotyping alleles.

likely to be the most common type of allele in the population, whereas the fraction of alleles that differ between individuals is likely to be those with a minor allele frequency in the population or genetic drift mutations. The same pattern was also observed with heterozygous indels, indicating that mutations may be biased to DNA loss.

Additional mutation and selection analyses done comparing the YH and NCBI36 genomes are available as Supplementary Information.

Structural variation identification

We used paired-end alignment information to identify structural variations. We identified structural variation boundaries between the YH and NCBI36 genomes by detecting abnormally aligned read pairs that have improper orientation relationships or span sizes (see Methods for details). We identified a total of 2,682 structural variations (Fig. 3a). Because our YH genome sequencing methodology generates paired-end reads with short but very accurate insert sizes, we could identify variations larger than 100 bp, about 6 times the insert size standard deviation. Identified structural variations had a median length of 492 bp, smaller than that of the database of genomic variants (DGV; 30.8 kb)[8]. This indicates that our methods were biased towards the detection of small structural variation events, but also indicates that it has an acceptable resolution as compared to current structural variation analyses[9–11].

Using paired-end methods, we identified more deletion (2,441) than duplication (33) events. Greater detection of deletions may be because they are identified by observing unexpectedly long insert sizes in paired-end clusters, whereas detection of insertions longer than our paired-end library span size will probably be missed.

We searched for candidate regions where larger insertions might have occurred by adopting a method based on the ratio of single-end to paired-end read depth and found 4,819 regions with a ratio significantly higher ($P < 0.001$) than the average ratio over the whole genome. Our data indicated that 4,377 (90.8%) of these candidate regions were likely to have insertions of repetitive elements, such as mammalian interspersed repeats (MIR; 2,067) and Alu elements (692) in the short interspersed nuclear elements (SINE) category, or L1 elements (1,601) in the long interspersed nuclear elements (LINE) category (see Methods for details).

Recent studies[10,11] have shown that novel sequences (those not anchored to the NCBI reference genome) are a considerable source of structural variations. To search for sequences unique to the YH genome, we analysed 487 million unmapped short reads. Among these, 0.39% could be aligned on unanchored scaffolds of NCBI36, 1.09% on novel small contigs of the Venter genome, and 0.67% on novel sequences identified by ref. 10. Using the *de novo* assembler
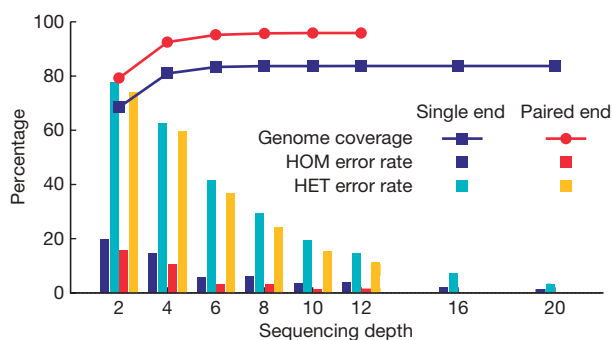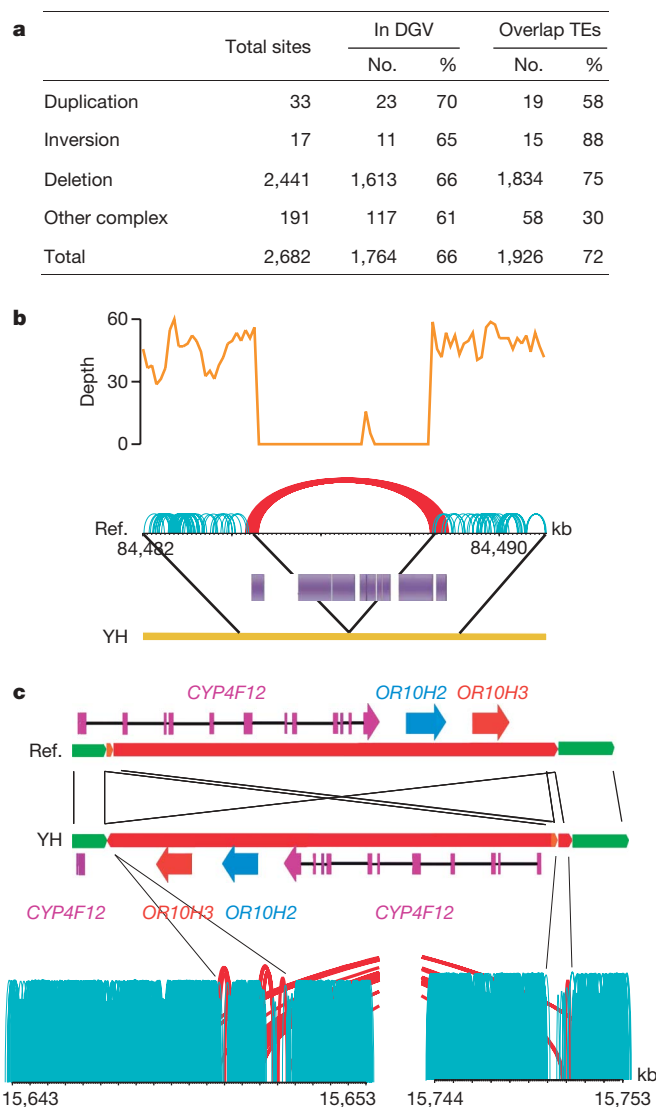
a

	Total sites	In DGV		Overlap TEs	
		No.	%	No.	%
Duplication	33	23	70	19	58
Inversion	17	11	65	15	88
Deletion	2,441	1,613	66	1,834	75
Other complex	191	117	61	58	30
Total	2,682	1,764	66	1,926	72

b

c

Figure 3 | Summary of structural variations. a, Abundance of each class of structural variation. The overlap with known structural variations in the DGV (http://projects.tcag.ca/variation/) and with transposons (transposable elements, TEs) was calculated. About 34% of our identified structural variations are novel (having less than 10% of a portion of the YH structural variations overlapping with structural variations in the DGV). Transposable elements are a major component of the identified deletions, with Alus and LINEs involved in 49% and 34% of the deletions, respectively. **b,** An example of a deletion of a transposon complex on YH chromosome 1. The sequencing depth by both single-end and paired-end reads are shown. Normally aligned paired-end reads are shown in green, whereas abnormally aligned paired-end reads, which have unexpected long insert sizes or an incorrect orientation relationship, are shown in red. **c,** An example of an inversion on YH chromosome 19. Local assembly showed that a 102,405-bp fragment was inverted and reinserted in the genome. There are three genes in this sequence fragment, and the last exon of gene *CYP4F12* was destroyed by this inversion event.

Velvet[12], we could assemble only 1,731,355 (0.36%) reads into 20,949 contigs with lengths >100 bp. In total, 10,398 (49.6%) of these contigs aligned well with unplaced human clones in GenBank. Of the remaining short contigs, 961 (4.6%) aligned with chimpanzee and mouse genomes at greater than 90% identity. These may represent deletions present in populations of European descent or be regions missed in the assembly of both NCBI36 and the Venter genome.

Because most structural variations occur in transposable elements or repetitive sequences, they are unlikely to have any major impact on function. (See Fig. 3b for an example of a deletion of a transposable

element complex.) In the YH genome, we did find structural variations that resulted in the complete or partial deletion of 33 genes, and 30.3% of these are homozygous deletions, increasing their likelihood of affecting gene function (Supplementary Table 6). An example of a gene disruption event is in the *CYP4F12* gene on YH chromosome 19, where an inversion has broken the gene into two segments (Fig. 3c). We used PCR amplification and sequencing to validate the inversion breakpoints. This gene also had non-synonymous mutations in its obsolete exons, indicating that it may have been under neutral selection.

Haplotype analysis

We used PHASE[13] and the available phased genotypes of the HapMap CHB/JPT population to predict the YH genome haplotypes. The 700,320 YH genome heterozygotes that overlapped with HapMap loci were used to construct 4,399 haplotype blocks that averaged 587 kb in size (Fig. 4). Of these heterozygous SNPs, 3,039 (0.43%) showed an inconsistent phase in the two adjacent fragments, which may potentially break the haplotype blocks. Additional potential haplotype breakpoints were 1,021,953 heterozygous YH genome SNPs absent in the HapMap. We evaluated this by checking paired-end reads that simultaneously covered two of the heterozygotes used in phasing. A total of 43,902 heterozygous SNP pairs were covered by read pairs, among which 97.37% (42,746 pairs) were in agreement with haplotypes as the corresponding covered read pairs. In total, the 2,434 haplotypes that had sizes greater than 200 kb covered 2.38 Gb of the genome.

Genetic ancestry

To estimate the ancestral composition of the YH individual's genome, we did a cluster analysis using an evenly sampled 87,614 loci with known alleles in all 270 HapMap individuals (Supplementary Fig. 3). The YH individual was estimated to share alleles[14] (thus ancestry) at 94.12% with the Asian, 4.12% with the European and 1.76% with the African populations. Collection of more data from all representative worldwide populations and development of analytical models to provide better estimates of time since admixture will improve the ability to assess an individual's personal genetic history.

Effective population size, N_e, is the number of breeding individuals in an idealized population that would show the same amount of allele frequency dispersion under random genetic drift or the same amount of inbreeding as the population under consideration[15]. Assuming an infinite-site model of neutral mutations and equilibrium of mutation and drift, and adopting the mutation rate used by ref. 16 with 2.63×10^{-8} per site per generation, we estimated that the effective Chinese population size is about 5,700. The same analysis based on the population mutation parameters of the YH, Watson, Venter, and NCBI36 genomes gives an estimate of 3,300 for the effective human population size, which is closer to the estimation based on HapMap
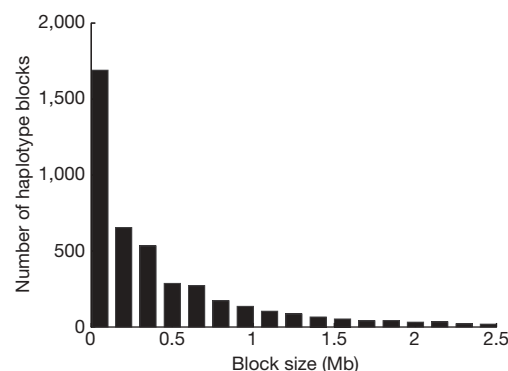
Figure 4 | Size distribution of predicted haplotype blocks of autosomes. Haplotypes were constructed using PHASE software with the 700,300 autosomal heterozygous SNPs that overlapped with the CHB/JPT genotypes from the HapMap phase II data.

data[17], but lower than the estimated 10,000–15,000 ancestral population size.

Known phenotypic or disease risk variant screen

The primary goal of personal genome sequencing is to allow identification of disease risk genotypes. We surveyed 1,495 alleles of 116 genes in the YH genome in the Online Mendelian Inheritance in Man (OMIM)[18] database and found one mutation in the *GJB2* gene, which is associated with a recessive deafness disorder. This allele was heterozygous, thus there was no expectation of, or evidence for, deafness in this individual, but it does raise the possibility of offspring having this disorder.

A preliminary search of genes and variants associated with common, complex phenotypes or disorders using OMIM data (Table 3) identified several genotypes that confer risk for tobacco addiction and Alzheimer's disease. This donor is a heavy smoker, as is consistent with individuals of similar genotypes in tobacco addiction studies. The donor contains 9 (56.3%) of the 16 identified Alzheimer's disease risk alleles[3], including two *APOE* alleles[19] and 7 *SORL1* alleles[20]. These findings indicate an increased risk for Alzheimer's disease, but there are no available data from any family members to assess whether there is a family history of Alzheimer's disease.

Discussion

Here we present the first genome sequence of an Asian individual. This sequence, which was accomplished using next-generation short-read sequencing technology, is one of the first genome sequences from a single individual (the genome sequences of J. D. Watson and J. C. Venter were accomplished using 454 and Sanger sequencing technology, respectively).

Our analysis of the YH genome, including consensus assembly, assessment of genome coverage, variation detection and validation, demonstrated the ability of this technology for sequencing large eukaryotic genomes given the availability of a reference genome. This sequencing method also resulted in sequence redundancy reaching an average 36-fold; significantly deeper than the ~7-fold coverage of the Watson and Venter genomes. Thus, the YH consensus sequence accuracy is higher and is especially suitable for calling heterozygous alleles.

Next-generation sequencing technologies have a very high throughput, as a hundred million DNA fragments can be sequenced in parallel on the chip. The Illumina GA sequencing used in this study can provide up to 4–8 Gb high-quality data per week. In this regard, the time needed to decipher a human genome (1–2 months using five next-generation sequencers), as well as the cost of sequencing (less than half a million US dollars), are substantially reduced.

The use of paired-end sequencing for structural variation detection allowed the identification of small but accurate insert sizes, making the attainable resolution excellent for deletion and small insertion identification, but limited for detection of insertions longer

than the paired-end insert sizes. Using a combination of both short and long insert sizes in the future will enable the identification of a larger variety of structural variations.

We were also able to phase a large number of heterozygous SNPs that overlapped with sites of inferred haplotypes of the Asian population from the HapMap data. However, to phase all the heterozygous SNPs of the assembled diploid genome with two sites covered by two reads belonging to a pair, we require different sized, long paired-end sequences. Improvement in haplotype prediction and heterozygote phasing will require genome sequences from many individuals in a population.

Adding to such advances, a recently formed international collaborative project, called the 1,000 Genome Project, aims to catalogue a detailed set of human genetic variations, which will serve as a multiple-genome-sequence blueprint for building genetic maps and extend our knowledge on genetic difference between individuals and between different ethnic populations. Ultimately, we predict an increase in the number of people who will be able to afford having their own genomes sequenced. Personal genome sequencing may eventually become an essential tool for diagnosis, prevention and therapy of human diseases.

METHODS SUMMARY

Library preparation followed the manufacturer's instructions (Illumina). Cluster generation was performed using the Illumina cluster station and the workflow was as follows: template hybridization, isothermal amplification, linearization, blocking, denaturation and sequencing primer hybridization. The fluorescent images were processed to sequences using the Illumina base-calling pipeline (SolexaPipeline-0.2.2.6). The human reference genome, together with the annotation of genes and repeats, were downloaded from the UCSC database (http://genome.ucsc.edu/), in line with NCBI build 36.1. dbSNP v128 and HapMap release 23 were used. The SNP set of the Venter genome was downloaded from the public FTP of JCVI, and the SNP set of the Watson genome was provided by Baylor College of Medicine.

We used SOAP to align all short reads onto the human reference genome (NCBI 36), and we used a statistical model based on Bayesian theory and the Illumina quality system to calculate the probability of each possible genotype at every position from the alignment of short reads on the NCBI reference genome. The genotype of each position was assigned as the allele types that had the highest probability. The final consensus probabilities were transformed to quality scores in Phred scale. We grouped abnormally mapped paired-end reads with coordinate distances smaller than the maximum insert size on both ends into diagnostic paired-end (PE) clusters. To avoid misalignment, PE clusters with <4 pairs were discarded. Common structural variations such as deletions, translocations, duplications, inversions and so on were examined and summarized into alignment models. The reads were assembled locally to verify the specific coordinate of structural variation elements.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. International Human Genome Sequencing Consortium.. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
4. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
5. Church, G. M. The personal genome project. *Mol. Syst. Biol.* **1**, doi:10.1038/msb4100040 (2005).
6. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
7. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
8. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
9. Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
10. Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
11. Bovee, D. *et al.* Closing gaps in the human genome with fosmid resources generated from multiple individuals. *Nature Genet.* **40**, 96–101 (2008).

Table 3 | Number of alleles identified that increase the risk to specific complex diseases

Traits	Associated genes	Associated SNPs	Predisposing alleles in YH	
			Number	Per cent
Alzheimer's	7	16	9	56.3
Diabetes	26	46	7	15.2
Hypertension	8	10	1	10.0
Obesity	6	27	1	3.7
Parkinson's	7	11	1	9.1
Hypolactasia	1	2	0	0.0
Alcohol addiction	3	3	0	0.0
Tobacco addiction	7	19	12	63.2

The genes and SNPs associated with complex diseases were from curated data sources. The results here are limited with regard to the conclusions that can be drawn, as nearly all of the SNPs associated with disease have been tested only in a relatively small number of samples, and haven't been tested in the Asian population.

12. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).

13. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).

14. Tang, H., Peng, J., Wang, P. & Risch, N. J. Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* **28**, 289–301 (2005).

15. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).

16. Noonan, J. P. *et al.* Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118 (2006).

17. Tenesa, A. *et al.* Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* **17**, 520–526 (2007).

18. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).

19. Coon, K. D. *et al.* A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J. Clin. Psychiatry* **68**, 613–618 (2007).

20. Rogaeva, E. *et al.* The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nature Genet.* **39**, 168–177 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions Ju.W., W.W., R.L. and Yi.L. contributed equally to this work. Ju.W., H.Y. and Ji.W. managed the project. W.W., G.T., Jun.Z., Jua.Z., Ya.L., Hu.L., Yi.Z., Zhenz.Y., X.Y., B.Y., W.L., Da.L., Q.H., Yu.L., S.Y., F.C., L.L., K.Z., Y.R., L.Y., Y.G., G.Y., Zhu.L., Xiaol.F., K.K., L.B. and X.Z. performed sequencing. Ju.W., R.L. and Yi.L. designed the analyses. R.L., Yi.L., W.F., J.L., Y.G., B.F., He.L., Xiaod.F., Z.D., Dong Li, Y.H., H.Z., I.H., M.I., J.P., Jin.Z., J.D., Ya.Z., J.Q., L.M., G.L., Zhent.Y., G.Z., C.Y., F.L, S.L, P.N., J.R., Q.L., Hongm.Z., Dongy.L, Zhi.L., N.L., G.G., Jia.Z., J.Y., L.F., Q.C., Y.S., A S., C.P., Hongk.Z., G.W., R.N., R.D. and S.L. performed the data analyses. Ju.W., R.L., Yi.L. and L.G. wrote the paper.

METHODS

DNA library construction and sequencing. Genomic DNA was extracted from peripheral venous blood, and the blood sample was collected using the guidelines dictated by the institutional review board of the Beijing Genomics Institute (BGI).

Library preparation followed the manufacturer's instructions (Illumina). Briefly, 2–5 μg of genomic DNA in 50 μl TE buffer were fragmented by nebulization with compressed nitrogen gas at 32 p.s.i. for 9 min. Nebulization generated double-stranded DNA fragments with blunt ends or with 3′ or 5′ overhangs. The overhangs were converted to blunt ends using T4 DNA polymerase and Klenow polymerase, after which an 'A' base was added to the ends of double-stranded DNA using Klenow exo− (3′ to 5′ exo minus). Next, DNA adaptors (Illumina) with a single 'T' base overhang at the 3′ end were ligated to the above products. These products were then separated on a 2% agarose gel, excised from the gel at a position between 150 and 250 bp, and purified (Qiagen Gel Extraction Kit). The adaptor-modified DNA fragments were enriched by PCR with PCR primers 1.1 and 2.1 (Illumina). Separate 8-, 10-, 12-, 15- and 18-cycle reactions were used for sequencing. The concentration of the libraries was measured by absorbance at 260 nm.

The template DNA fragments of the constructed libraries were hybridized to the surface of flow cells and amplified to form clusters. After double-stranded DNA was denatured to single-stranded DNA and nonspecific sites were blocked, genomic DNA sequencing primers were hybridized for DNA sequencing initiation. In brief, cluster generation was performed on the Illumina cluster station, and the basic workflow (based on the standard Illumina protocol) was as follows: template hybridization, isothermal amplification, linearization, blocking and denaturisation, and hybridization of the sequencing primers. The fluorescent images were converted to sequence using the Illumina base-calling pipeline (SolexaPipeline-0.2.2.6).

Public data used. The human reference genome, together with genes and repeats annotation, was downloaded from the UCSC database (http://genome.ucsc.edu/), which has the same sequence as the NCBI reference build 36.1. The NCBI reference genes with prefix 'NM' were mapped to the reference genome using BLAT by UCSC. Hits with >90% identity were retained for further analysis, and only one transcript was retained for each gene. dbSNP v128 and HapMap release 23 were used. The SNP set from the Venter genome was downloaded from the public FTP site of JCVI (ftp://ftp.jcvi.org/pub/data/huref/), and the SNP set of the Watson genome was provided by Baylor College of Medicine.

Short reads alignment. We used SOAP to align each read or read-pair to a position on a chromosome of the NCBI36 human reference genome that had least number of nucleotide differences between the read and the reference genome, and called this a 'best hit'. If a read had only a single best hit, it was considered uniquely aligned. Reads that had more than one 'best hit' (meaning they could be aligned to multiple positions that each had the same number of mismatches) were considered repeatedly aligned. For repeatedly aligned reads a random position was chosen from all of its best hits for placement on the reference genome for sequencing depth calculation.

In the specific alignment process, at most two mismatches were allowed between a read and the reference, and best hits were selected. Because errors can accumulate during sequencing, the quality of the last several base pairs at the end of reads can be relatively low. We thus set option −c 52 during our alignment procedure. Thus, if a read could not be aligned, we discarded the first base, and iteratively trimmed 2 bp at the 3′ end until the read could be aligned or the remaining sequence was shorter than 27 bp. For paired-end reads, two reads belonging to a pair were aligned with both being in the correct orientation and proper span size on the reference genome. If a pair could not be aligned without gaps but allowing at most two mismatches on each read, a gapped alignment was then performed with a maximum gap size of 3 bp. If the two reads could not be aligned as a pair, they were aligned independently.

Consensus assembly. We used a statistical model based on Bayesian theory and the Illumina quality system to calculate the probability of each possible genotype at every position from the alignment of short reads on the NCBI reference genome. A calibration matrix was built based on all uniquely mapped reads to estimate the probability for a given genotype T to have an observed base X located at a position k of its original read with quality score S. For a variety of reasons, similar sequencing errors are often repeated, thus, the ith occurrence of base X covering a particular position would contribute less to denote an X in consensus by an adjustment formula. In brief, likelihood $P(X|T)$ is a function of (k, S, i, X, T), not simply of $F(S)$. The total likelihood of all observed bases (O) covering a site $P(O|T)$ is the product of each one.

From observed prior probability, the SNP rate is expected to be about 0.1%, and the most common SNPs should already be present in dbSNP. Therefore, for positions without known polymorphisms, on one haploid, the reference bases will dominate the prior probability as 0.999; others will share the remaining 0.1% mutation rate. Because sequencing errors would look like heterozygous (HET) SNPs, a penalty factor of 0.001 is multiplied to the HET prior probability. At dbSNP sites, bases already observed dominate the prior probability equally and the HET penalty factor is 0.01. As a result, the prior probabilities were as follows: (1) 0.45 for a homozygote and 0.1 for a heterozygote at a SNP site that has been validated in dbSNP; (2) 0.495 for a homozygote and 0.01 for a heterozygote at a SNP site that has not been validated in dbSNP; and (3) 1×10^{-6} for a homozygote and 2×10^{-6} for a heterozygote at a potentially novel SNP site (one that is absent in dbSNP).

Using the information above, we calculated the posterior probability of each genotype using a Bayesian formula. The genotype of each position was assigned as the allele type that had the highest probability. A rank sum test was applied to adjust for the probability of heterozygotes. The final consensus probabilities were transformed to quality scores in Phred scale.

Calling SNPs. We used six steps to filter out unreliable portions of the consensus sequence: (1) we used a Q20 quality cutoff; (2) we required at least four reads; (3) the overall depth, including randomly placed repetitive hits, had to be less than 100; (4) the approximate copy number of flanking sequences had to be less than 2 (this was done to avoid misreading SNPs as heterozygotes caused by the alignment of similar reads from repeat units or by copy number variations (CNVs)); (5) there had to be at least one paired-end read; and (6) the SNPs had to be at least 5 bp away from each other. For chromosome X and Y, condition (2) was altered by requiring only two unique reads with at least 1 paired-end (PE) read. In the SOAP algorithm, a gap-free alignment is done first and then a gapped alignment. Thus, we required condition (6) because most of the discrepancies between the YH genome and the NCBI reference genome that are too close to each other are due to mismatches across indels. After filtering, we were confident in the calculated YH consensus sequence, and discrepancies between the YH genome and NCBI reference genome were called as SNPs.

Identification of short indels. As the number of SNPs is roughly one order of magnitude larger than that of indels, in the first stage of alignment we did not allow any gaps. Thus, some read pairs containing real indels could not be mapped when PE requirements were satisfied. After the first alignment stage, we mapped the unmapped read pairs by allowing up to 3-bp indels to enable them to meet PE requirements. This limited the indels that could be detected in our study to gaps of 1–3 bp in length. If different read pairs provided the same outer coordinates in mapping, they are likely to be duplicated products of a single fragment during PCR. We merged these redundant pairs before looking for indels. Gaps that were supported by at least three non-redundant paired-end reads were extracted. If the number of ungapped reads that crossed a possible indel was no more than twice that of gapped reads, then an indel was called. In chromosome X and Y, we required all indel sites to be covered by only gapped reads because valid indels on sex chromosomes are expected to be homozygous.

ARTICLES

DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

Timothy J. Ley[1,2,3,4]*, Elaine R. Mardis[2,3]*, Li Ding[2,3], Bob Fulton[3], Michael D. McLellan[3], Ken Chen[3], David Dooling[3], Brian H. Dunford-Shore[3], Sean McGrath[3], Matthew Hickenbotham[3], Lisa Cook[3], Rachel Abbott[3], David E. Larson[3], Dan C. Koboldt[3], Craig Pohl[3], Scott Smith[3], Amy Hawkins[3], Scott Abbott[3], Devin Locke[3], LaDeana W. Hillier[3,8], Tracie Miner[3], Lucinda Fulton[3], Vincent Magrini[2,3], Todd Wylie[3], Jarret Glasscock[3], Joshua Conyers[3], Nathan Sander[3], Xiaoqi Shi[3], John R. Osborne[3], Patrick Minx[3], David Gordon[8], Asif Chinwalla[3], Yu Zhao[1], Rhonda E. Ries[1], Jacqueline E. Payton[5], Peter Westervelt[1,4], Michael H. Tomasson[1,4], Mark Watson[3,4,5], Jack Baty[6], Jennifer Ivanovich[4,7], Sharon Heath[1,4], William D. Shannon[1,4], Rakesh Nagarajan[4,5], Matthew J. Walter[1,4], Daniel C. Link[1,4], Timothy A. Graubert[1,4], John F. DiPersio[1,4] & Richard K. Wilson[2,3,4]

Acute myeloid leukaemia is a highly malignant haematopoietic tumour that affects about 13,000 adults in the United States each year. The treatment of this disease has changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered. Whole-genome sequencing is now possible at a reasonable cost and timeframe to use this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes. Here we present the results obtained from sequencing a typical acute myeloid leukaemia genome, and its matched normal counterpart obtained from the same patient's skin. We discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse, the function of which is not yet known. Our study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.

We used massively parallel sequencing technology to sequence the genomic DNA of tumour and normal skin cells obtained from a patient with a typical presentation of French–American–British (FAB) subtype M1 acute myeloid leukaemia (AML) with normal cytogenetics. For the tumour genome, 32.7-fold 'haploid' coverage (98 billion bases) was obtained, and 13.9-fold coverage (41.8 billion bases) was obtained for the normal skin sample. Of the 2,647,695 well-supported single nucleotide variants (SNVs) found in the tumour genome, 2,584,418 (97.6%) were also detected in the patient's skin genome, limiting the number of variants that required further study. For the purposes of this initial study, we restricted our downstream analysis to the coding sequences of annotated genes: we found only eight heterozygous, non-synonymous somatic SNVs in the entire genome. All were new, including mutations in protocadherin/cadherin family members (*CDH24* and *PCLKC* (also known as *PCDH24*)), G-protein-coupled receptors (*GPR123* and *EBI2* (also known as *GPR183*)), a protein phosphatase (*PTPRT*), a potential guanine nucleotide exchange factor (*KNDC1*), a peptide/drug transporter (*SLC15A1*) and a glutamate receptor gene (*GRINL1B*). We also detected previously described, recurrent somatic insertions in the *FLT3* and *NPM1* genes. On the basis of deep readcount data, we determined that all of these mutations (except *FLT3*) were present in nearly all tumour cells at presentation and again at relapse 11 months later, suggesting that the patient had a single dominant clone containing all of the mutations. These results demonstrate the power of whole-genome sequencing to discover new cancer-associated mutations.

AML refers to a group of clonal haematopoietic malignancies that predominantly affect middle-aged and elderly adults. An estimated 13,000 people will develop AML in the United States in 2008, and 8,800 will die from it[1]. Although the life expectancy from this disease has increased slowly over the past decade, the improvement is predominantly because of improvements in supportive care—not in the drugs or approaches used to treat patients.

For most patients with a 'sporadic' presentation of AML, it is not yet clear whether inherited susceptibility alleles have a role in the pathogenesis[2]. Furthermore, the nature of the initiating or progression mutations is for the most part unknown[3]. Recent attempts to identify additional progression mutations by extensively re-sequencing tyrosine kinase genes yielded very few previously unidentified mutations, and most were not recurrent[4,5]. Expression profiling studies have yielded signatures that correlate with specific cytogenetic subtypes of AML, but have not yet suggested new initiating mutations[6–8]. Recent studies using array-based comparative genomic hybridization and/or single nucleotide polymorphism (SNP) arrays, although identifying important gene mutations in acute lymphoblastic leukaemia[9,10] have revealed very few recurrent submicroscopic somatic copy number variants in AML (M.J.W., manuscript in preparation, and refs 11–13). Together, these studies suggest that we have not yet discovered most of the relevant mutations that contribute to the pathogenesis of AML. We therefore believe that unbiased whole-genome sequencing will be required to identify most of these mutations. Until recently, this approach has not been feasible because of the high cost of conventional

[1]Department of Medicine, [2]Department of Genetics, [3]The Genome Center at Washington University, [4]Siteman Cancer Center, [5]Department of Pathology and Immunology, [6]Division of Biostatistics, and [7]Department of Surgery, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [8]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.
*These authors contributed equally to this work.

capillary-based approaches and the large numbers of primary tumour cells required to yield the necessary genomic DNA. 'Next-generation' sequencing approaches, however, have changed this landscape.

Our group has pioneered the use of whole-genome re-sequencing and variant discovery approaches using the Illumina/Solexa technology with the genome of the nematode worm *Caenorhabditis elegans* as a proof-of-principle[14]. This approach has distinct advantages in reduced cost, a markedly increased data production rate, and a low input requirement of DNA for library construction. In the present study, we used a similar approach to sequence the tumour genome of a single AML patient and the matched normal genome (derived from a skin biopsy) of the same patient. After alignment to the human reference genome, sequence variants were discovered in the tumour genome and compared to the patient's normal sequence, to the dbSNP database, and to variants recently reported for two other human genomes[15,16]; revealing new single nucleotide and small insertion/deletion (indel) variants genome-wide. Somatic mutations were detected in genes not previously implicated in AML pathogenesis, demonstrating the need for unbiased whole-genome approaches to discover all mutations associated with cancer pathogenesis.

Rationale for using the FAB M1 AML subtype for sequencing

Of the eight FAB subtypes of AML, M1 AML is one of the most common (~20% of all cases). No specific cytogenetic abnormalities or somatic initiating mutations have been identified for this subtype; in fact, about half of the patients with *de novo* M1 AML have normal cytogenetics[17–19]. The frequency of well-described progression mutations (for example, activating alleles of *FLT3*, *KIT* and *RAS*) is similar to that of other common FAB subtypes[5]. We therefore decided to sequence the genome of tumour cells derived from a patient with M1 AML, because so little is known about the molecular pathogenesis of this common subtype. The criteria used to select the sample are outlined in Supplementary Information.

Case presentation of UPN 933124

The case presentation is described in detail in the Supplementary Information. In brief, a previously healthy woman in her mid-50s presented suddenly with fatigue and easy bruisability, and was found to have a peripheral white blood cell count of 105,000 cells per microlitre, with 85% myeloblasts. A bone marrow examination revealed 100% myeloblasts with morphological features and cell surface markers consistent with FAB M1 AML (Supplementary Fig. 1). Cytogenetic analysis of tumour cells revealed a normal 46,XX karyotype. Although the patient experienced a complete remission with conventional therapies, she relapsed at 11 months and expired 24 months after her initial diagnosis was made. At relapse, the bone marrow had 78% myeloblasts, and contained a new clonal cytogenetic abnormality, t(10; 12) (p12; p13). Informed consent for whole-genome sequencing was subsequently obtained from her next of kin.

A typical M1 AML diploid genome and expression profile

The tumour sample from patient 933124 contained no somatic copy number changes at a resolution of ~5 kb (further confirmed on the NimbleGen 2.1M array platform, data not shown), and no evidence of copy number neutral loss-of-heterozygosity (LOH), indicating that the genome was essentially diploid at this level of resolution (see Supplementary Fig. 2). Further analysis of the 933124-derived tumour and skin samples showed 26 inherited copy number variants (that is, detected in both the tumour and skin samples). All but two of these had been previously reported in the Database of Genomic Variants (see Supplementary Table 1). All of the copy number variants detected in this genome were found in at least one other AML patient (89 other cases, mostly Caucasian, have been queried using the same SNP array platform), and all but one were found in at least one of the 160 Caucasian HapMap and Coriell samples that were studied on the same array platform (Supplementary Table 1).

To determine whether the tumour cells of 933124 were typical of M1 AML, we compared the expression signatures of 111 *de novo* AML cases using unsupervised clustering (Ward's method, see Supplementary Information). The expression profile of patient 933124 clustered with multiple other M1 (and M2) AML cases with normal cytogenetics, suggesting that the genetic events underlying the pathogenesis of this case are similar to those of other cases exhibiting normal cytogenetics (Supplementary Fig. 3).

Coverage depth of the tumour and skin genomes

Because most of the acquired mutations in cancer genomes have been shown to be heterozygous, the complete sequencing of a cancer genome requires the detection of both alleles at most positions in the genome[20]. We therefore designed sequence coverage metrics to define the point at which 90% diploid coverage had been reached. To minimize errors associated with any single platform or measurement, diploid coverage for this genome was assessed using a set of high-quality SNPs derived from two different SNP array platforms, Affymetrix 6.0 and Illumina Infinium 550K. For a SNP to be included in the high-quality set, the following criteria had to be satisfied: (1) identical genotypes were called from both assays at the same genomic positions, and (2) the resulting genotype was heterozygous. For the 933124 tumour genome, 46,494 heterozygous SNPs passed the above criteria and were defined as high-quality SNPs. For the skin samples, 46,572 high-quality SNPs were defined.

We performed 98 full runs on the Illumina Genome Analyser to achieve the targeted level of 90% diploid coverage as determined by coverage of the high-quality SNP set. Maq[21] was used to perform alignment, determine consensus, and identify SNVs within the 98 billion bases generated from the tumour genome (see Table 1). Maq predicted a total of 3.81 million SNVs (Maq SNP quality $\geq$ 15) in the tumour genome, including matching heterozygous genotypes for 91.2% of the 46,494 high-quality SNPs. When we lowered the Maq SNP quality cutoff to 0, 94.06% high-quality SNPs were predicted. Further investigation of Maq alignments revealed coverage for both alleles at a further 5.38% of the high-quality SNPs, but Maq did not predict a SNP or matching heterozygous genotype owing to insufficient depth or quality of coverage. Extra analysis revealed coverage at 46,484 of 46,494 high-quality SNPs for at least one allele (that is, 99.98% haploid coverage for the tumour genome).

We sequenced the genome of normal skin cells from the same patient to enable the identification of inherited sequence variants in the tumour genome. Our targeted diploid coverage goal for the skin-derived genome was 80%. We achieved this goal with only 34 Solexa runs (41.8 billion bases), using improved reagents and longer read lengths to attain 82.6% diploid and 84.2% haploid coverage (Table 1).

To begin evaluating the quantity and quality of the detected sequence variants in the tumour and skin genomes, we compared the overlap and uniqueness of this genome's variants with respect to the James D. Watson and J. Craig Venter genomes, and to dbSNP (v127; Fig. 1). Of the 3.68 million single nucleotide variants (SNVs; Maq SNP quality $\geq$15, excluding SNVs found on chromosome X) predicted by Maq in the tumour genome, 2.36 million were present in dbSNP, 2.36 million were detected in the skin genome (Fig. 1a), 1.50 million were detected in the Venter genome, and 1.58 million were found in the Watson genome (Fig. 1b). Ultimately, 1.70 million SNVs were unique to the 933124 tumour genome. On filtering the 933124 SNVs at different Maq quality values to determine the stability of results, we observed that the proportion of 933124 SNVs that also are in dbSNP increases from 63.9% to 69.48% when the Maq quality threshold score increases from 15 to 30, as expected.

Refining the detection of potential somatic mutations

Because the number of sequence variants initially detected by Maq was high, we developed improved filtering tools to effectively separate true variants from false positives. To this end, we generated an

Table 1 | Tumour and skin genome coverage from patient 933124

	Tumour	Skin
Libraries	4	3
Runs	98	34
Reads obtained	5,858,992,064	2,122,836,148
Reads passing quality filter	3,025,923,365	1,228,177,690
Bases passing quality filter	98,184,511,523	41,783,794,834
Reads aligned by Maq	2,729,957,053	1,080,576,680
Reads unaligned by Maq	295,966,312	138,276,594
SNVs detected with respect to hg18 (no Y)	3,811,115	2,918,446
SNVs (chr 1–22) detected with respect to hg18	3,681,968 (100.0%)	2,830,292 (100.0%)
SNVs also present in dbSNP	2,368,458 (64.3%)	2,161,695 (76.4%)
SNVs also present in Venter genome	1,499,010 (40.7%)	1,383,431 (48.9%)
SNVs also present in Watson genome	1,573,435 (42.7%)	1,456,822 (51.5%)
SNVs not in dbSNP/Venter/Watson	1,223,830 (33.2%)	591,131 (20.9%)
SNVs not in dbSNP/Venter/Watson/skin	925,200 (25.1%)	—
HQ SNPs	46,494 (100.0%)	46,572 (100.0%)
HQ SNPs where reference allele is detected	42,419 (91.2%)	38,454 (82.6%)
HQ SNPs where variant allele is detected	43,164 (92.9%)	39,220 (84.2%)
HQ SNPs where both alleles are detected	42,415 (91.2%)	38,454 (82.6%)

Assessments are shown of the haploid and diploid coverage of the tumour and skin genomes from AML patient 933124.
Chr, chromosome; hg18, human genome version 18; HQ, high quality.

experimental data set by re-sequencing Maq-predicted SNVs, randomly selecting a training subset and a test data set, whose annotations and features were submitted to Decision Tree C4.5 (ref. 22).

This approach identified parameters that separated true variants from false positives, revealing that SNV-supporting read counts (unique on the basis of read start position and base position in supporting reads), base quality and Maq quality scores are chief determinants for identifying false positives. Implementing rules obtained from the Decision Tree analysis resulted in 91.9% sensitivity and 83.5% specificity for validated SNVs.

Identification of somatic mutations in coding sequences

The patient had 3,813,205 sequence variants in her tumour genome, as defined by Maq scores of >15 (Table 1). Of these, 2,647,695 were supported by the Decision Tree analysis in the tumour genome, of which 2,584,418 (97.6%) were also detected in the skin genome (Fig. 2). The detailed algorithm for selecting putative somatic variants is described in Supplementary Information. Most of the 63,277 tumour-specific variants we detected were either present in dbSNP or were previously described in the Watson or Venter genomes (31,645), or occurred in non-genic regions (20,440). A total of 11,192 variants were located within the boundaries of annotated
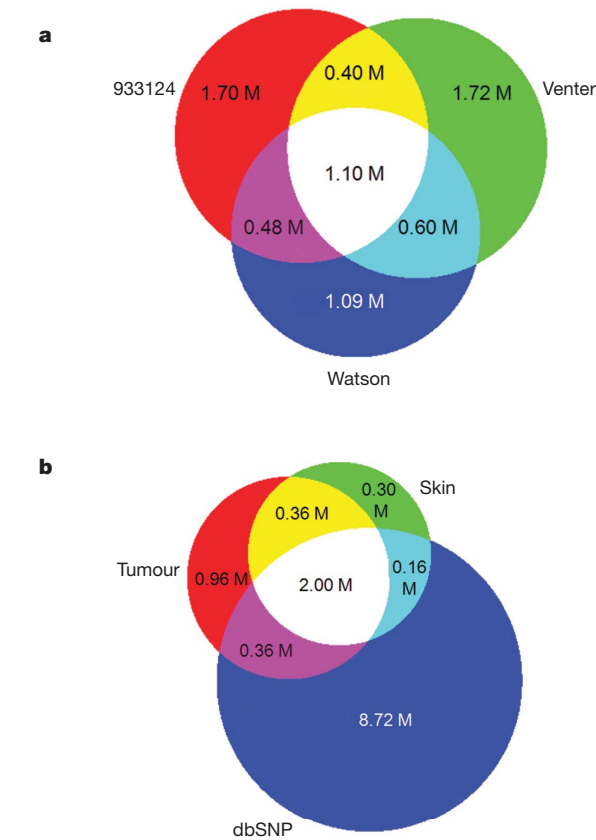
Figure 1 | Overlap of SNPs detected in 933124 and other genomes. a, Venn diagram of the overlap between SNPs detected in the 933124 tumour genome and the genomes of J. D. Watson and J. C. Venter. **b,** Venn Diagram of the overlap among the 933124 tumour genome, the skin genome and dbSNP (ver. 127). SNVs were defined with a Maq SNP quality ≥15.
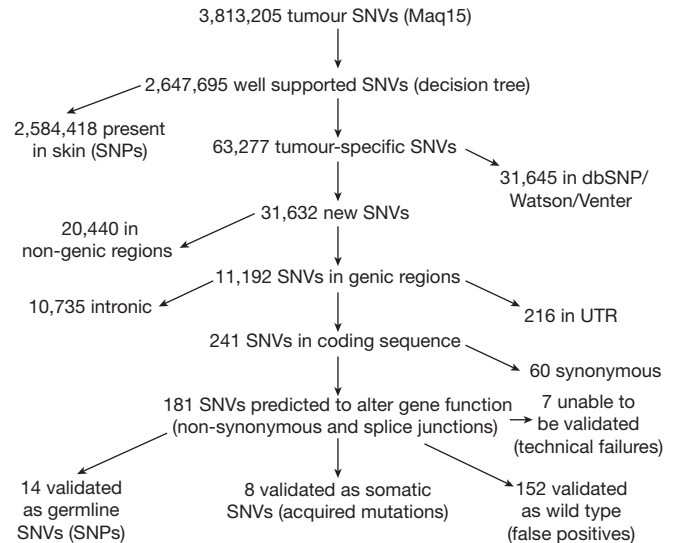
Figure 2 | Filters used to identify somatic point mutations in the tumour genome. See text for details. UTR, untranslated regions.

genes; 216 of these variants were in untranslated regions, and 10,735 were in introns (but not involving splice junctions) and were not explored further in our analysis. Of the coding sequence variants, 60 were synonymous, and not further evaluated. The remaining 181 variants were either non-synonymous, or were predicted to alter splice site function. By sequencing polymerase chain reaction (PCR)-generated amplicons from the tumour and skin samples (and also from the relapse tumour sample obtained 11 months after the original presentation), we determined that 152 of these variants were false positive (that is, wild type) calls, 14 were inherited SNPs, and eight were somatic mutations in both the original tumour and the relapse sample (Table 2). Seven variants could not be validated, either because the regions involved were repetitive, or because all attempts to obtain PCR amplicons failed. All of the PCR-amplified exons from the eight genes containing validated somatic mutations were sequenced in 187 further cases of AML using samples from our discovery and validation sets[23]; no further somatic mutations were detected in these genes (data not shown). A description of how we estimated the false negative (12.45%) and false positive (0.06%) rates for SNVs over the entire genome is presented in Supplementary Information. Using these estimates, we can predict that very few somatic, non-synonymous variants were missed by our analysis of this deeply covered genome.

Defining mutation frequencies in the tumour sample

To better define the percentage of tumour cells that contained each of the discovered somatic mutations, we amplified each mutation-containing locus from non-amplified genomic DNA derived from the *de novo* and relapse tumour samples, and from the skin biopsy obtained at presentation. The resulting amplicons were sequenced using the Roche/454 FLX platform, and the frequency of reads containing the reference and variant alleles were defined (Fig. 3 and Table 3). Control amplicons containing a known heterozygous SNP in *BRCA2* (encoding N372H) and a homozygous SNP in *TP53* (encoding P72R) were analysed similarly. The *BRCA2* SNP yielded ~50% variant frequencies in the tumour and skin samples, whereas nearly 100% of the *TP53* alleles were variant in all three samples, as expected. Remarkably, all eight somatic SNVs were detected at ~50% frequencies in the primary tumour sample (100% blasts), and at ~40% frequencies in the relapse sample (78% blasts; if the variant frequencies are corrected for blast counts—that is, multiplied by 1.28—the frequencies at relapse also were ~50%). The NPMc (cytoplasmic nucleophosmin) mutation was also detected at a frequency of ~50%, but the *FLT3* internal tandem duplication (ITD) allele was only detected in 35.1% of the 454 reads at diagnosis and 31.3% at relapse, suggesting that the mutation was not present in all tumour cells at diagnosis or relapse.

Notably, the variant alleles also were detected at frequencies of ~5–13% in the skin sample. In retrospect, it is clear that the skin sample contained contaminating leukaemic cells, because the patient's white blood cell count at presentation was 105,000 per microlitre, with 85% blasts. This information was used to inform the Decision Tree analysis described above: we allowed high-quality
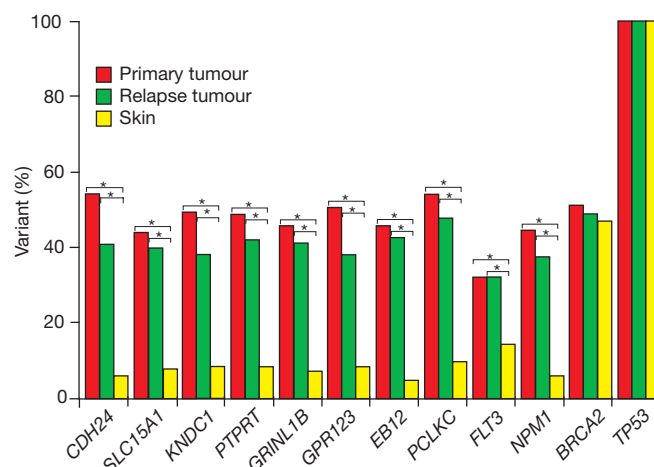
Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that of the skin sample for all mutations ($P < 0.000001$ for all mutations, Fisher's exact test, denoted by a single asterisk in all cases). Note that the normal skin sample was contaminated with leukaemic cells containing the somatic mutations. The patient's white blood cell count was 105,000 (85% blasts) when the skin punch biopsy was obtained.

tumour variants to move forward in the discovery pipeline if they were detected at a low frequency (two or fewer reads) in the skin sample, as defined by a binomial test.

Detecting insertions and deletions (indels)

To discover small indels (<6 bp) from sequence reads (32–35 bp long), we started with a set of 236 million reads that were not confidently aligned by Maq to the reference genome. We applied Cross_Match and BLAT to identify gapped alignments that are unique in the genome. To detect indels longer than 6 bp, we developed a 'split reads' algorithm (see Supplementary Information) that aligns sub-segments of reads independently to the genome, and computes a mapping quality for the derived gapped alignment on the basis of the number of hits and the quality of the bases. These efforts resulted in the identification of 726 putative small indels (1 to 30 bp in size) that occur in coding exons, 393 of which (54.2%) were found in dbSNP. After manual review, we selected a set of 28 putative somatic coding indels for validation using PCR-based dye terminator sequencing. Of these putative indels, 22 were validated but were found present in both tumour and skin (15 of these were in dbSNP), two were false positive calls, two had no coverage, and two were previously validated somatic insertions in *NPM1* (4 bp) and *FLT3* (30 bp).

Discussion

Here we describe the sequencing and analysis of a primary human cancer genome using next-generation sequencing technology. Our

Table 2 | Non-synonymous somatic mutations detected in the AML sample

Gene	Consequence	Type	Solexa tumour reads WT:variant	Solexa skin reads WT:variant	Conservation score of mutant base	Mutations in other AML cases*
CDH24	Y590X	Nonsense	9:9	16:0	0.998	0/187
SLC15A1	W77X	Nonsense	15:12	19:0	1.000	0/187
KNDC1	L799F	Missense	7:8	20:0	NA	0/187
PTPRT	P1235L	Missense	9:13	16:0	1.000	0/187
GRINL1B	R176H	Missense	15:10	14:0	NA	0/187
GPR123	T38I	Missense	11:11	13:0	NA	0/187
EBI2	A338V	Missense	7:12	18:2	1.000	0/187
PCLKC	P1004L	Missense	19:9	15:1	0.98	0/187
FLT3	ITD	Indel	18:12	8:0	NA	51/185
NPM1	CATG ins	Indel	36:6	33:0	NA	43/180

Ins, insertion; WT, wild type.
* Patient cohort defined in ref. 23.

Table 3 | 454 Readcount data for somatic mutations and known SNPs

Gene	Consequence	Primary AML (100% blasts)			Skin			Relapse (78% blasts)		
		Variant	Ref	Variant (%)	Variant	Ref	Variant (%)	Variant	Ref	Variant (%)
CDH24	Y590X	5672	4890	53.70	564	10358	5.16	3108	4599	40.33
SLC15A1	W77X	3817	4962	43.48	875	10773	7.51	4714	7173	39.66
KNDC1	L799F	4640	4848	48.90	770	8972	7.90	3883	6342	37.98
PTPRT	P1235L	998	1058	48.54	126	1489	7.80	350	493	41.52
GRINL1B	R176H	2211	2674	45.26	318	4461	6.65	1447	2070	41.14
GPR123	T38I	4618	4569	50.27	850	9751	8.02	3660	6057	37.67
EBI2	A338V	12750	15453	45.21	458	10088	4.34	2646	3627	42.18
PCLKC	P1004L	992	855	53.71	341	3153	9.76	705	773	47.70
FLT3	ITD	4220	7810	35.08	3475	23159	13.05	3870	8495	31.30
NPM1	CATG ins	1550	1974	43.98	143	2390	5.65	2303	3910	37.07
BRCA2	N372H	778	752	50.85	763	876	46.55	285	303	48.47
TP53	P72R	8989	1	99.99	8161	0	100.00	7914	6	99.92

The differences between variant frequencies in primary or relapse tumour samples and skin were highly significant for all somatic mutations ($P < 0.000001$, Fisher's exact test, one tailed). The *BRCA2* variant is a known heterozygous SNP in this genome, and the *TP53* variant is a known homozygous SNP.

patient's tumour genome was essentially diploid, and contained ten non-synonymous somatic mutations that may be relevant for her disease. These mutations affect genes participating in several well-described pathways that are known to contribute to cancer pathogenesis, but most of these genes would not have been candidates for directed re-sequencing on the basis of our current understanding of cancer. Hence, these results justify the use of next-generation whole-genome sequencing approaches to reveal somatic mutations in cancer genomes.

As we demonstrated in our re-sequencing of the genome of the *C. elegans* N2 Bristol strain[14], and again in this study, massively parallel short-read sequencing provides an effective method for examining single nucleotide and short indel variants by comparison of the aligned reads to a reference genome sequence. By sequencing our patient's tumour genome to a depth of >30-fold coverage, and gauging our ability to detect known heterozygous positions across the genome, we have produced a sufficient depth and breadth of sequence coverage to comprehensively discover somatic genome variants. A slightly lower coverage of the normal genome from this individual helped to identify nearly 98% of potential variants as being inherited, a critical filter that allowed us to more readily identify the true somatic mutations in this tumour. Our results strongly support the notion that hypothesis-driven (for example, candidate gene-based) examination of tumour genomes by PCR-directed or capture-based methods is inherently limited, and will miss key mutations. A further and important consideration is the demand for large amounts of genomic DNA by these techniques; this is a serious limitation when precious clinical samples are being studied. The Illumina/Solexa technology requires only ~1 μg of DNA per library, enabling the study of primary tumour DNA rather than requiring the use of tumour cell lines, which may contain genetic changes and adaptations required for immortalization and maintenance in tissue culture conditions.

A total of ten non-synonymous somatic mutations were identified in this patient's tumour genome. Two are well-known AML-associated mutations, including an internal tandem duplication of the *FLT3* receptor tyrosine kinase gene, which constitutively activates kinase signalling, and portends a poor prognosis[5,24,25], and a four-base insertion in exon 12 of the *NPM1* gene (NPMc)[26–28]. Both of these mutations are common (25–30%) in AML tumours, and are thought to contribute to progression of the disease rather than to cause it directly[29]. Notably, the frequency of the mutant *FLT3* allele in the primary and relapse tumour samples (35.08% and 31.30%, respectively) was significantly less than that of the other nine mutations ($P < 0.000001$ for both the primary and relapse samples). These data suggest that the *FLT3* ITD may not have been present in all tumour cells, and further, that it may have been the last mutation acquired.

The other eight somatic mutations that we detected are all single base changes, and none has previously been detected in an AML genome. Four of the genes affected, however, are in gene families that are strongly associated with cancer pathogenesis (including

PTPRT, *CDH24*, *PCLKC* and *SLC15A1*). The other four somatic mutations occurred in genes not previously implicated in cancer pathogenesis, but whose potential functions in metabolic pathways suggest mechanisms by which they could act to promote cancer (including *KNDC1*, *GPR123*, *EBI2* and *GRINL1B*). We speculate about the roles of these mutations for the pathogenesis of this patient's disease in Supplementary Information.

The importance of the eight newly defined somatic mutations for AML pathogenesis is not yet known, and will require functional validation studies in tissue culture cells and mouse models to assess their relevance. Even though we could not detect recurrent mutations in the limited AML sample set that we surveyed, several lines of evidence suggest that these mutations may not be random, 'passenger' mutations. First, somatic mutations in this genome are extremely rare. The rarity of somatic variants, and the normal diploid structure of the tumour genome, argues strongly against genetic instability or DNA repair defects in this tumour. Conceptually, this result is further supported by the very small number of somatic mutations discovered in the expressed tyrosine kinases of AML samples[4,5]; genetic instability does not seem to be a general feature of AML genomes.

Second, on the basis of the equivalent frequencies of the variant and wild-type alleles for the mutations in the tumour genome (except for *FLT3* ITD), it is highly probable that all the mutations are heterozygous, and are present in virtually all of the tumour cells (Fig. 3). The latter suggests that these mutations may have all been selected for and retained because they are important for disease pathogenesis in this patient. Alternatively, all may have occurred simultaneously in the same leukaemia-initiating cell, but only a subset of the mutations (or an as-yet undetected mutation) is truly important for pathogenesis (that is, disease 'drivers' versus passengers). Although we suggest that the latter hypothesis is very unlikely on the basis of our current understanding of tumour progression, many more AML genomes will need to be sequenced to resolve this issue.

Third, the same mutations were detected in tumour cells in the relapse sample at approximately the same frequencies as in the primary sample. All of these mutations were therefore present in the resistant tumour cells that contributed to the patient's relapse, further suggesting that a single clone contains all ten mutations. Fourth, seven of the ten genes containing somatic mutations were detectably expressed in the tumour sample. *FLT3* and *NPM1* messenger RNAs were highly expressed in this tumour sample, as they are in virtually all AML samples. We detected mRNA from the *CDH24*, *SLC15A1* and *EBI2* genes on the Affymetrix expression array, whereas expression of *GRINL1B* and *PCLKC* were detected by PCR with reverse transcription (RT–PCR; data not shown). Expression of *KNDC1*, *PTPRT* and *GPR123* was not detected by either approach, but we cannot rule out expression of these genes in a small subset of tumour cells (for example, leukaemia-initiating cells). Furthermore, for the five point mutations where data are available, the mutated base is highly conserved across multiple species (Table 2).

Although we performed whole-genome sequencing on this cancer sample, we restricted our initial validation studies to the 1–2% of the genome that encodes genes. This raises the issue of whether sequencing the complementary DNA transcriptome of this tumour would have been a faster, cheaper and more efficient way of finding the mutations. Although this approach will undoubtedly be an important adjunct to whole-genome sequencing, there are several advantages to the approach we used: (1) coverage models for whole-genome libraries are at present better understood than for cDNA libraries, where transcript abundance can vary over many orders of magnitude; (2) even if the transcriptome had been sequenced, extensive characterization of the normal genome would have been required to distinguish inherited variants from somatic mutations; and (3) relevant non-synonymous mutations could be missed by cDNA sequencing, including mutations that result in RNA instability (splice variants, nonsense mutations), and/or mutations in genes expressed at low levels, or in only a small subset of tumour cells.

The additional non-coding and non-genic somatic variants in this genome (which we presently estimate at 500–1,000 on the basis of our calculated false positive and negative rates for non-synonymous mutations), will provide a rich source of potentially relevant sequence changes that will be better understood as more cancer genomes are sequenced.

In summary, we have successfully used a next-generation whole-genome sequencing approach to identify new candidate genes that may be relevant for AML pathogenesis. We cannot overemphasize the importance of parallel sequencing of the patient's normal genome to determine which variants were inherited; the identification of the true somatic mutations in this tumour genome would not have been feasible without this approach. Furthermore, until hundreds (or perhaps thousands) of normal genomes and other AML tumours are sequenced, the contextual relevance of the mutations found in this genome will be unknown. Nevertheless, the somatic mutations that we did find were neither predicted by the curation of previously defined cancer genes, nor by the study of this tumour using unbiased, high-resolution array-based genomic approaches. For AML and other types of cancer, whole-genome sequencing may therefore be the only effective means for discovering all of the mutations that are relevant for pathogenesis.

METHODS SUMMARY

Sequence end reads (average length for tumour genome, 32 bp, and for skin, 35 bp) were generated from Illumina/Solexa fragment libraries derived from the tumour or skin cells of patient 933124, using the Illumina Genome Analyser. The analysed reads were aligned to the human reference genome (NCBI Build 36) using Maq[21]. Coverage of the tumour and normal genomes was ascertained by comparison to the patient's heterozygous SNPs, established by compiling shared SNP calls monitored on the Affymetrix 6.0 and Illumina Infinium 550K genotyping platforms. We examined the Maq alignments by Decision Tree analysis to discover SNVs, as well as to identify copy number variants. Non-aligned reads were further analysed for indel discovery. For all putative variants, we attempted validation using custom PCR and capillary sequencing on the ABI 3730 platform. All validated somatic mutations were further analysed by Roche/454 sequencing of PCR-generated amplicons made from primary genomic DNA to compare readcounts of wild-type and mutant alleles in the primary tumour, skin and relapse tumour samples. A complete description of the AML case sequenced, and the materials and methods used to generate this data set are provided in the Supplementary Information.

Sequence variant deposition in dbGaP. High-quality sequence variants defined by Decision Tree (2,647,695 variants) will be deposited in the dbGaP database (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap) for review by approved investigators.

1. Jemal, A. *et al.* Cancer statistics, 2008. *CA Cancer J. Clin.* **58**, 71–96 (2008).
2. Owen, C., Barnett, M. & Fitzgibbon, J. Familial myelodysplasia and acute myeloid leukaemia—a review. *Br. J. Haematol.* **140**, 123–132 (2008).
3. Mrozek, K., Marcucci, G., Paschka, P., Whitman, S. P. & Bloomfield, C. D. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* **109**, 431–448 (2007).
4. Loriaux, M. M. *et al.* High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. *Blood* **111**, 4788–4796 (2008).
5. Tomasson, M. H. *et al.* Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. *Blood* **111**, 4797–4808 (2008).
6. Schoch, C. *et al.* Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc. Natl Acad. Sci. USA* **99**, 10008–10013 (2002).
7. Bullinger, L. *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1605–1616 (2004).
8. Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1617–1628 (2004).
9. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
10. Mullighan, C. G. *et al.* BCR–ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110–114 (2008).
11. Raghavan, M. *et al.* Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res.* **65**, 375–378 (2005).
12. Paulsson, K. *et al.* High-resolution genome-wide array-based comparative genome hybridization reveals cryptic chromosome changes in AML and MDS cases with trisomy 8 as the sole cytogenetic aberration. *Leukemia* **20**, 840–846 (2006).
13. Rucker, F. G. *et al.* Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J. Clin. Oncol.* **24**, 3887–3894 (2006).
14. Hillier, L. W. *et al.* Whole-genome sequencing and variant discovery in *C. elegans. Nat. Methods* **5**, 183–188 (2008).
15. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
16. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
17. Byrd, J. C. *et al.* Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* **100**, 4325–4336 (2002).
18. Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322–2333 (1998).
19. Mrozek, K., Heerema, N. A. & Bloomfield, C. D. Cytogenetics in acute leukemia. *Blood Rev.* **18**, 115–136 (2004).
20. Wendl, M. C. & Wilson, R. K. Aspects of coverage in medical DNA sequencing. *BMC Bioinformatics* **9**, 239 (2008).
21. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* doi:10.1101/gr.078212.108 (in the press).
22. Quinlan, J. R. *C4.5:Programs for Machine Learning* 302 (Morgan Kaufmann Publishers, 1993).
23. Link, D. C. *et al.* Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. *Blood* **110**, 1648–1655 (2007).
24. Frohling, S. *et al.* Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* **12**, 501–513 (2007).
25. Levis, M. & Small, D. FLT3: ITDoes matter in leukemia. *Leukemia* **17**, 1738–1752 (2003).
26. Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* **352**, 254–266 (2005).
27. Thiede, C. *et al.* Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood* **107**, 4011–4020 (2006).
28. den Besten, W., Kuo, M. L., Williams, R. T. & Sherr, C. J. Myeloid leukemia-associated nucleophosmin mutants perturb p53-dependent and independent activities of the Arf tumor suppressor protein. *Cell Cycle* **4**, 1593–1598 (2005).
29. Kelly, L. M. *et al.* PML/RARα and FLT3-ITD induce an APL-like disease in a mouse model. *Proc. Natl Acad. Sci. USA* **99**, 8283–8288 (2002).

LETTERS

Prospects for detecting supersymmetric dark matter in the Galactic halo

V. Springel[1], S. D. M. White[1], C. S. Frenk[2], J. F. Navarro[3,4], A. Jenkins[2], M. Vogelsberger[1], J. Wang[1], A. Ludlow[3] & A. Helmi[5]

Dark matter is the dominant form of matter in the Universe, but its nature is unknown. It is plausibly an elementary particle, perhaps the lightest supersymmetric partner of known particle species[1]. In this case, annihilation of dark matter in the halo of the Milky Way should produce γ-rays at a level that may soon be observable[2,3]. Previous work has argued that the annihilation signal will be dominated by emission from very small clumps[4,5] (perhaps smaller even than the Earth), which would be most easily detected where they cluster together in the dark matter haloes of dwarf satellite galaxies[6]. Here we report that such small-scale structure will, in fact, have a negligible impact on dark matter detectability. Rather, the dominant and probably most easily detectable signal will be produced by diffuse dark matter in the main halo of the Milky Way[7,8]. If the main halo is strongly detected, then small dark matter clumps should also be visible, but may well contain no stars, thereby confirming a key prediction of the cold dark matter model.

If small-scale clumping and spatial variations in the background are neglected, then it is easy to show that the main halo would be much more easily detected than the haloes of known satellite galaxies. For a smooth halo of given radial profile shape, for example that given in ref. 9 by Navarro, Frenk and White (NFW), the annihilation luminosity can be written as $L \propto V_{max}^4/r_{half}$, where V_{max} is the maximum of the circular velocity curve and r_{half} is the radius containing half the annihilation flux. (For an NFW profile $r_{half} = 0.089 r_{max}$, where r_{max} is the radius at which the circular velocity curve peaks.) The flux from an object at distance d therefore scales as $V_{max}^4/(r_{half} d^2)$, whereas the angular size of the emitting region scales as r_{half}/d. Hence, the signal-to-noise ratio for detection against a bright uniform background scales as $S/N \propto C V_{max}^4/(r_{half}^2 d)$. The constant C depends only weakly on profile shape (see Supplementary Information). For the cold dark matter (CDM) simulation of the Milky Way's halo we present below, $V_{max} \approx 209$ km s^{-1}, $r_{max} \approx 28.4$ kpc and $d \approx 8$ kpc. Using parameters for Milky Way (MW) satellite haloes from previous modelling[10,11], the highest S/N is predicted for the Large Magellanic Cloud (LMC), for which $V_{max} \approx 65$ km s^{-1}, $r_{max} \approx 13$ kpc and $d = 48$ kpc, leading to $(S/N)_{MW}/(S/N)_{LMC} = 134$. (Note that this overestimates the contrast achievable in practice; see Supplementary Information.)

The simulations used in this Letter are part of the Virgo Consortium's Aquarius Project[12] to simulate the formation of CDM haloes similar to that of the Milky Way. The largest simulation has a dark matter particle mass of $1,712 M_\odot$ (where $M_\odot$ is the solar mass) and a converged length scale of 120 pc, both of which improve by a factor of three on the largest previous simulation[13]. This particular halo has mass $M_{200} = 1.84 \times 10^{12} M_\odot$ within $r_{200} = 246$ kpc, the radius enclosing a mean density 200 times the critical value.

Simulations of the same object at mass resolutions lower by factors of 8, 28.68, 229.4 and 1,835 enable us to check explicitly for the convergence of the various numerical quantities presented below.

The detectable annihilation luminosity density at each point within a simulation is

$$\mathcal{L}(\mathbf{x}) = \mathcal{G}(\text{particle physics, observational set-up}) \rho^2(\mathbf{x})$$

where $\rho(\mathbf{x})$ is the local dark matter density and the constant $\mathcal{G}$ does not depend on the structure of the system but encapsulates the properties of the dark matter particle (for example, annihilation cross-section and branching ratio into photons) as well as those of the telescope and observation. For the purposes of this Letter, we set $\mathcal{G} = 1$ and give results only for the relative luminosities and detectability of the different structures. In this way, we can quote results that are independent of the particle physics model and the observational details.

Figure 1 shows the distribution of annihilation radiation within our Milky Way halo as a function of the resolution used to simulate it. This plot excludes the contribution to the emission from resolved substructures. Half of the emission from the Milky Way halo is predicted to come from within 2.57 kpc and 95% from within 27.3 kpc. For the lowest resolution simulation (1,835 times coarser than the largest simulation), the luminosity is clearly depressed below 3 kpc, but for the second-best simulation, it converges well for $r > 300$ pc, where r is the distance to the centre. Thus we infer that the largest simulation should give convergent results to $r \approx 150$ pc, and that numerical resolution affects the luminosity of the main diffuse halo only at the few per cent level. Note that much larger effects will be caused by the baryonic component of the Milky Way, which we neglect. This is expected to compress the inner dark matter distribution and thus to enhance its annihilation signal[14], which would strengthen our conclusions. (See Supplementary Information for discussion of this and related topics.)

Within 433 kpc of the halo centre, we identify 297,791 and 45,024 self-bound subhaloes in our two highest resolution simulations. Many of these can be matched individually in the two simulations, allowing a crucial (and not previously attempted) test of the convergence of their internal structure. In Fig. 2 we show the results of such a test. The values inferred for V_{max} show no systematic offsets between simulation pairs down to the smallest objects detected in the lower resolution simulation, suggesting that V_{max} values are reliable above ~ 1.5 km s^{-1} in the largest simulation. Systematic offsets are visible in each simulation at small r_{max}, reaching 10% on a scale that decreases systematically as the resolution increases. From this, we conclude that our largest simulation produces r_{max} values that are accurate to 10% for $r_{max} > 165$ pc. Figure 1 shows that almost all the annihilation signal from a halo comes from $r \ll r_{max}$, corresponding to scales that

---

[1]Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, 85740 Garching, Germany. [2]Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK. [3]Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia V8P 5C2, Canada. [4]Department of Astronomy, University of Massachusetts, Amherst, Massachusetts 01003-9305, USA. [5]Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands.
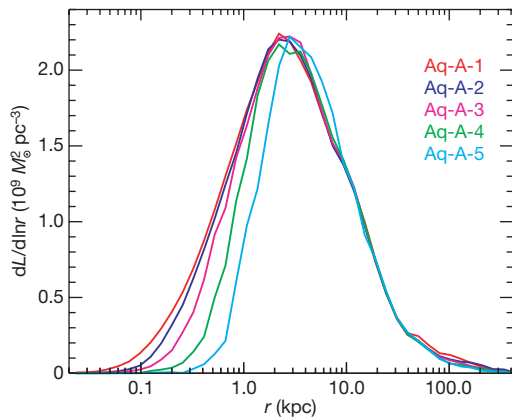
73

Figure 1 | Annihilation luminosity as a function of radius for the diffuse dark matter component of Milky Way haloes. These simulations assume $\mathcal{G} = 1$ and a Universe with mean matter density $\Omega_m = 0.25$, cosmological constant density $\Omega_\Lambda = 0.75$, Hubble constant $H_0 = 73\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, primordial spectral index $n = 1$ and present fluctuation amplitude $\sigma_8 = 0.9$. In this representation, the total emitted luminosity is proportional to the area under each curve. The particle mass (in units of $M_\odot$) in the simulations is 1,712 for simulation Aq-A-1, and grows to 1.37×10^4, 4.91×10^4, 3.93×10^5 and 3.14×10^6 for simulations Aq-A-2, Aq-A-3, Aq-A-4 and Aq-A-5, respectively. The fluctuations at large radii are due to subhaloes below our detection limit. These curves were calculated by estimating a density local to each N-body particle through a Voronoi tesselation of the full particle distribution, and then summing the annihilation luminosities of individual particles in a set of logarithmically spaced spherical shells. Note that the vertical axis is linear, so these curves demonstrate numerical convergence at the per cent level in the detailed structure of our main halo down to scales below 1 kpc.

are not well resolved for most subhaloes. In the following we will therefore assume the annihilation luminosity from the diffuse component of each subhalo to be $L = 1.23 \mathcal{G} V_{max}^4/G^2 r_{max}$, the value expected for an object with NFW structure (here G is the gravitational constant).

When estimating the Milky Way's annihilation luminosity from our simulations, we need to include the following components: (1) smooth emission associated with the main halo (hereafter, MainSm); (2) smooth emission associated with resolved subhaloes (SubSm); (3) emission associated with unresolved substructure in the main halo (MainUn); and (4) emission associated with substructure within the subhaloes themselves (SubSub). (Here we do not discuss emission from dark matter caustics[15].) These four components have very different radial distributions, both within the Milky Way and within its substructures. Neglect of this crucial fact in previous work (see below) has led to incorrect assessments of the importance of small-scale substructure for the detectability of the annihilation radiation.

The solid blue line in Fig. 3 shows $M(<r)/M_{200}$, where $M(<r)$ is the mass within r. Half of M_{200} lies within 98.5 kpc and only 3.3% within the solar circle ($r = 8$ kpc). The solid red line shows the corresponding curve for the MainSm annihilation luminosity, normalized by L_{200}, its value at r_{200}. This component is much more centrally concentrated than the mass; its half-luminosity radius is only 2.62 kpc. In contrast, the thick green line shows that the SubSm luminosity is much less centrally concentrated than the mass. This is a result of the dynamical disruption of substructure in the inner regions of the halo. The thick green line includes contributions from all substructures with mass exceeding $10^5 M_\odot$, almost all of which have converged values for V_{max} and r_{max}. This line is also normalized by L_{200}. Within r_{200}, SubSm contributes 76% as much luminosity as MainSm, but within 30 kpc, for example, this fraction is only 2.5%. The three thin green lines in Fig. 3 show the results of excluding contributions from less massive subhaloes, corresponding to thresholds $M_{thr} = 10^6 M_\odot$, $10^7 M_\odot$ and $10^8 M_\odot$. These all have similar shape and are offset approximately equally in amplitude, implying that SubSm luminosity
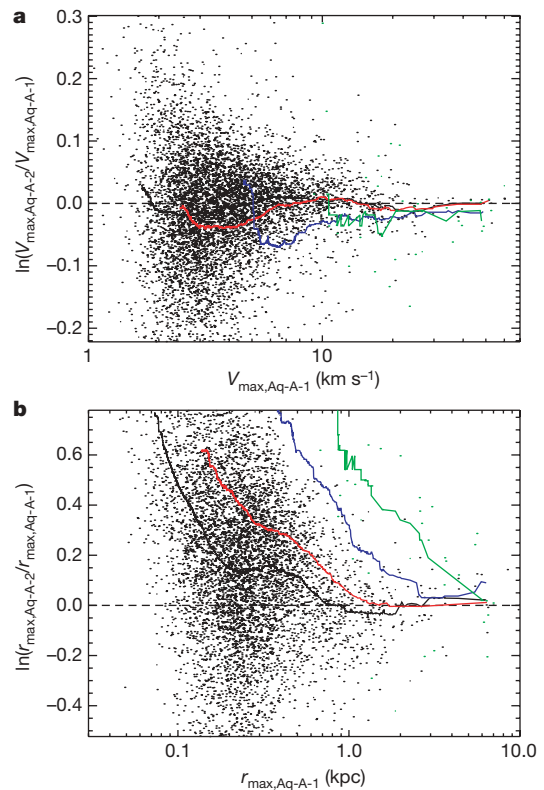

Figure 2 | Structural properties of dark matter subhaloes as a function of simulation resolution. a, $\ln(V_{max,Aq-A-2}/V_{max,Aq-A-1})$ against $V_{max,Aq-A-1}$ for 6,711 matched subhaloes detected by the SUBFIND algorithm[30] within 433 kpc of halo centre in our two highest resolution simulations, Aq-A-1 and Aq-A-2. The radius 433 kpc encloses an overdensity 200 times the cosmic mean. The black solid line shows the running median of this distribution. Red, blue and green lines give similar median curves for matches of the lower resolution simulation, Aq-A-1. **b,** As above, but for the ratio of characteristic sizes (r_{max}) as a function of that in the highest resolution simulation. We have checked that convergence in the subhalo mass is similarly good and that these results apply equally well to subhaloes inside 50 kpc.

scales as $M_{thr}^{-0.226}$ at all radii. If we assume, in the absence of other information, that this behaviour continues down to a minimum mass of $10^{-6} M_\odot$, which might be appropriate if the dark matter is the lightest supersymmetric particle[16], then MainUn and SubSm have the same radial distribution. We predict these two components together to be 232 times more luminous than MainSm within r_{200}, but still only 7.8 times more luminous within 30 kpc. A distant observer would thus infer the substructure population of the Milky Way to be 232 times brighter than its smooth dark halo, but from the Earth's position the total boost is predicted to be only 1.9 as the substructure signal typically comes from much larger distances.

We must now consider the additional luminosity due to (sub-)substructures (SubSub). Before a subhalo is accreted onto the main object, we assume its detailed structure to be similar to that of the main halo (including its subhalo population), but scaled down appropriately in mass and radius. (We have checked that such a scaling does indeed hold approximately for small independent objects outside the main halo.) However, once the subhalo is accreted, its outer regions are rapidly removed by tidal stripping. The longer a subhalo has been part of the main system and the closer it is to the centre, the more drastic is the stripping[17,18]. As a result, most of the substructure associated with the subhalo is removed, whereas its smooth luminosity is little affected. The removed (sub-)subhaloes are, in effect, transferred to components SubSm and MainUn.

A subhalo at Galactocentric distance r is typically truncated at tidal radius $r_t = (M_{sub}/[(2 - d\ln M/d\ln r)M(<r)])^{1/3} r$. We estimate its
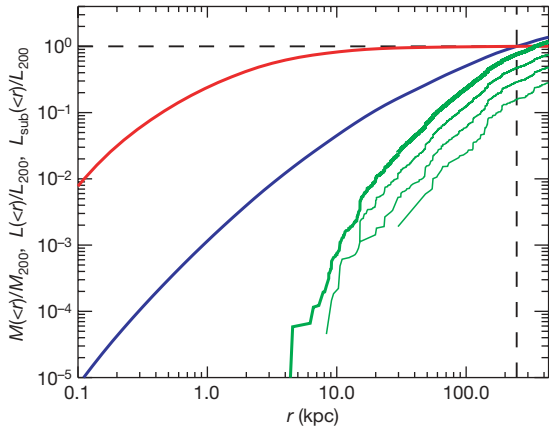
Figure 3 | Radial dependence of the enclosed mass and annihilation luminosity of various halo components. The blue line gives enclosed dark matter mass in units of M_{200}, the value at r_{200} (the radius enclosing a mean density 200 times the critical value, marked in the plot by a vertical dashed line). The red line gives the luminosity of smooth main halo annihilation (MainSm) in units of L_{200}, its value at r_{200}. The green lines give the luminosity of smooth subhalo annihilation (SubSm) for various lower limits to the subhalo mass considered; the solid thick line is for $M_{min} = 10^5 M_\odot$, the thin lines for $M_{min} = 10^6 M_\odot$, $10^7 M_\odot$ and $10^8 M_\odot$. Note that the shape of these lines is insensitive to M_{min}, and that their normalization is proportional to $M_{min}^{-0.226}$.

SubSub luminosity by assuming that all material beyond r_t is simply removed. The remaining SubSub luminosity can then be obtained from the curves of Fig. 3 if we scale them to match the measured parameters of the subhalo (M_{sub}, V_{max} and r_{max}). We assume that the r_{200} of the subhalo before accretion was proportional to its present V_{max}. (r_{200} is indeed nearly proportional to V_{max} for isolated haloes in our simulations.) We further assume that the ratio of subhalo mass to SubSub luminosity within r_t corresponds to the ratio between main halo mass and SubSm luminosity (from Fig. 3) within the scaled radius r_t/f, where $f = (V_{max}/209\,\mathrm{km\,s^{-1}})$. We must also correct for the SubSub luminosity below the mass limit $M_{min} = 10^5 f^3 M_\odot$, scaling down the resolution limit of our simulation appropriately for the subhalo. The SubSub luminosity must then be boosted by a factor of $(M_{min}/M_{lim})^{0.226}$, where M_{lim} is the free-streaming mass (which is $10^{-6} M_\odot$ in the example given above). For definiteness, we adopt $M_{lim} = 10^{-6} M_\odot$ in the discussion below, although none of our conclusions would change if we adopted, for example, $M_{lim} = 10^{-12} M_\odot$.

We now consider the expected appearance and detectability of these various components. The diffuse emission from the Milky Way's halo (MainSm) is distributed across the whole sky, falling away smoothly from the Galactic Centre. A randomly placed observer at $r = 8$ kpc would see half the flux within 13° of the Galactic Centre, most of this well outside the Galactic plane where contamination is strongest. Assuming NFW structure for individual subhaloes, half the diffuse emission from each object falls within the angular radius corresponding to $r_{half} = 0.089 r_{max}$. Because of their large typical distances, these subhaloes are almost uniformly distributed across the sky. The luminosity from unresolved subhaloes (MainUn) is similarly distributed and will appear smooth in γ-ray sky maps, with a centre to anticentre surface brightness contrast of only 1.54. Half the luminosity from (sub-)subhaloes within an individual subhalo falls within an angular radius corresponding to ~$0.6 r_t$; this is usually much more extended than the SubSm emission from the same subhalo.

This information allows us finally to calculate the relative detectability of the various components. As argued above, the signal-to-noise ratio for detection by an optimal filter against a bright uniform background can be written as $S/N \propto F/\theta_h$, where F is the total flux, θ_h is the angle containing half this flux, and the constant of proportionality depends weakly on profile shape but strongly on the particle physics

and observational parameters (the factor $\mathcal{G}$ above). To account for the finite angular resolution of the observation, we replace θ_h with $\theta_h' = (\theta_h^2 + \theta_{psf}^2)^{1/2}$. For example, $\theta_{psf} \approx 10$ arcmin is the characteristic point spread function of the LAT detector of the recently launched Fermi Gamma-ray Space Telescope (formerly GLAST) at the relevant energies[19]. In reality, the background at these energies is not uniform and is relatively poorly known[20,21]. In the Supplementary Information, we show that this is likely to reduce the detectability of the main smooth halo relative to that of subhaloes by a factor of up to ten in comparison with the numbers we quote below, which are based on the simple assumption of a uniform background.

In Fig. 4 we combine data for 1,000 randomly placed observers at $r = 8$ kpc. Figure 4a shows histograms of the S/N for detecting SubSm
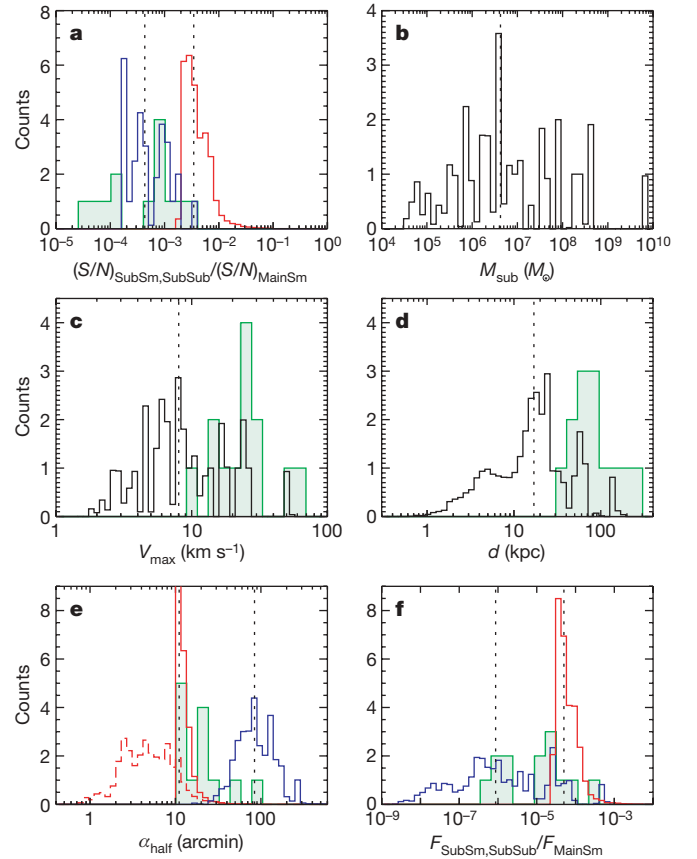

Figure 4 | Observability of subhaloes. Histograms of the properties of the 30 highest-S/N subhaloes that would be seen by each of 1,000 notional 'observers' placed at random 8 kpc from halo centre, assuming a 10-arcmin observational beam. The histograms are divided by 1,000 so they sum to 30. **a**, The 30 highest S/N values for SubSm (red) and SubSub (blue) emission. These do not necessarily come from the same subhaloes. The SubSub S/N values lie well below the SubSm values—subhalo detectability is not influenced by internal substructure. **b–d**, Histograms of the masses, V_{max} values and distances of the 30 haloes with highest-S/N SubSm emission. For these same haloes, **e** and **f** show half-light radii and fluxes separately for the SubSm (red) and the SubSub (blue) emission. In **e**, the dashed and solid red histograms show values before and after convolution with the telescope beam. For subhaloes with $V_{max} < 5$ km s^{-1} we have suppressed numerical noise by replacing the measured r_{max} by a value drawn from a suitably scaled version of the distribution measured at larger V_{max} for subhaloes within 50 kpc. This substitution has a modest effect on the low mass tails of our distributions. Fluxes are expressed in units of the flux from the main halo. Dashed vertical lines mark median values. The single highest-S/N subhaloes detected by each of our 1,000 'observers' are biased towards smaller and nearer objects; their median values are $(S/N)_{SubSm} = 0.015(S/N)_{MainSm}$, $V_{max} = 6$ km s^{-1}, $M_{sub} = 2 \times 10^6 M_\odot$ and $d = 4$ kpc. Light green histograms show the distributions predicted for SubSm emission from 13 known satellites of the Milky Way, based on published mass models[11].

and SubSub emission from the 30 highest-S/N subhaloes, and also shows the expected S/N for known satellites of the Milky Way. These are all expressed in units of the S/N for detecting the MainSm emission. Three important conclusions follow immediately: (1) no subhalo is expected to have S/N more than ~10% that of the main halo, even accounting for the expected effects of the non-uniform background; (2) the most easily detectable dark subhalo is predicted to have five times larger S/N than the LMC; and (3) the S/N predicted for SubSub emission is always much lower than that predicted for SubSm emission because of the much greater angular extent of the former.

Figure 4b–f shows histograms of the masses, V_{max} values, distances, angular half-light radii, and fluxes (relative to the flux from the main halo) of the 30 highest-S/N subhaloes. These are compared with the distributions for the known satellites of the Milky Way where appropriate. For the fluxes and half-light radii we show separate histograms for the SubSm and SubSub emission. A second set of conclusions follow. If subhaloes are detected, then the highest-S/N systems will (4) typically have masses and circular velocities well below those inferred for the currently known satellites of the Milky Way; (5) have angular half-light radii below 10 arcmin and so will not be resolved by Fermi; (6) be at distances ~4 kpc; and (7) typically have SubSm and SubSub fluxes that are factors of 10^{-4} and 10^{-6} times lower than those of the main halo, respectively.

These conclusions differ substantially from earlier work. Very small-scale substructure (below the resolution limit of our simulations) does not affect the detectability of dark matter annihilation in the Milky Way's halo. This is true both for the smooth main halo (contradicting refs 4, 5, 22, among others) and for its subhaloes (contradicting refs 6, 23, 24). Emission should be much more easily detected from the main halo than from subhaloes (contradicting refs 25, 26, but in agreement with ref. 27), even though the total flux is dominated by substructures (contradicting refs 28, 29). The most easily detectable subhalo is expected to be a relatively nearby object of lower mass than any known Milky Way satellite (contradicting refs 23, 25). Almost all of these differences stem from the differing spatial distribution of small-scale substructure and smooth dark matter, which our simulations are able to trace reliably because of their high resolution.

The Fermi satellite is now in orbit and accumulating a γ-ray image of the whole sky. If supersymmetry exists and the parameters of the theory are favourable, in a few years we may have a direct image of the Galaxy's dark halo. If we are really lucky, we may also detect substructures both without and with stars. This would provide a convincing confirmation of the CDM theory.

1. Bertone, G., Hooper, D. & Silk, J. Particle dark matter: Evidence, candidates and constraints. *Phys. Rep.* **405**, 279–390 (2005).
2. Gehrels, N. & Michelson, P. GLAST: The next-generation high energy gamma-ray astronomy mission. *Astropart. Phys.* **11**, 277–282 (1999).
3. Baltz, E. A. *et al.* Pre-launch estimates for GLAST sensitivity to dark matter annihilation signals. *J. Cosmol. Astropart. Phys.* **7**, 013, doi:10.1088/1475-7516/2008/07/013 (2008).
4. Calcáneo-Roldán, C. & Moore, B. Surface brightness of dark matter: Unique signatures of neutralino annihilation in the galactic halo. *Phys. Rev. D* **62**, 123005 (2000).
5. Berezinsky, V., Dokuchaev, V. & Eroshenko, Y. Small-scale clumps in the galactic halo and dark matter annihilation. *Phys. Rev. D* **68**, 103003 (2003).
6. Strigari, L. E., Koushiappas, S. M., Bullock, J. S. & Kaplinghat, M. Precise constraints on the dark matter content of Milky Way dwarf galaxies for gamma-ray experiments. *Phys. Rev. D* **75**, 083526 (2007).
7. Berezinsky, V., Bottino, A. & Mignola, G. High energy gamma-radiation from the Galactic center due to neutralino annihilation. *Phys. Lett. B* **325**, 136–142 (1994).
8. Bergström, L., Ullio, P. & Buckley, J. H. Observability of gamma rays from dark matter neutralino annihilations in the Milky Way halo. *Astropart. Phys.* **9**, 137–162 (1998).
9. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
10. Strigari, L. E. *et al.* Redefining the missing satellites problem. *Astrophys. J.* **669**, 676–683 (2007).
11. Peñarrubia, J., McConnachie, A. W. & Navarro, J. F. The cold dark matter halos of local group dwarf spheroidals. *Astrophys. J.* **672**, 904–913 (2008).
12. Springel, V. *et al.* in *High Performance Computing in Science and Engineering, Munich 2007* (eds Wagner, S., Steinmetz, M., Bode, A. & Brehm, M.) 93–108 (Springer, 2008).
13. Diemand, J. *et al.* Clumps and streams in the local dark matter distribution. *Nature* **754**, 735–738 (2008).
14. Prada, F., Klypin, A., Flix, J., Martínez, M. & Simonneau, E. Dark matter annihilation in the Milky Way galaxy: Effects of baryonic compression. *Phys. Rev. Lett.* **93**, 241301 (2004).
15. Mohayaee, R., Shandarin, S. & Silk, J. Dark matter caustics and the enhancement of self-annihilation flux. *J. Cosmol. Astropart. Phys.* **5**, 015, doi:10.1088/1475-7516/2007/05/015 (2007).
16. Hofmann, S., Schwarz, D. J. & Stöcker, H. Damping scales of neutralino cold dark matter. *Phys. Rev. D* **64**, 083507 (2001).
17. De Lucia, G. *et al.* Substructures in cold dark matter haloes. *Mon. Not. R. Astron. Soc.* **348**, 333–344 (2004).
18. Gao, L., White, S. D. M., Jenkins, A., Stoehr, F. & Springel, V. The subhalo populations of ΛCDM dark haloes. *Mon. Not. R. Astron. Soc.* **355**, 819–834 (2004).
19. Michelson, P. F. in *The First GLAST Symposium* (eds Ritz, S., Michelson, P. & Meegan, C. A.) 8–12 (AIP Conference Series, Vol. 921, American Institute of Physics, 2007).
20. Hunter, S. D. *et al.* EGRET observations of the diffuse gamma-ray emission from the galactic plane. *Astrophys. J.* **481**, 205–240 (1997).
21. Strong, A. W., Moskalenko, I. V. & Reimer, O. Diffuse galactic continuum gamma rays: A model compatible with EGRET data and cosmic-ray measurements. *Astrophys. J.* **613**, 962–976 (2004).
22. Aloisio, R., Blasi, P. & Olinto, A. V. Gamma-ray constraints on neutralino dark matter clumps in the galactic halo. *Astrophys. J.* **601**, 47–53 (2004).
23. Kuhlen, M., Diemand, J. & Madau, P. The dark matter annihilation signal from galactic substructure: Predictions for GLAST. *Astrophys. J.* (in the press); preprint at ⟨http://arxiv.org/abs/0805.4416⟩.
24. Strigari, L. E. *et al.* The most dark-matter-dominated galaxies: Predicted gamma-ray signals from the faintest Milky Way dwarfs. *Astrophys. J.* **678**, 614–620 (2008).
25. Evans, N. W., Ferrer, F. & Sarkar, S. A travel guide to the dark matter annihilation signal. *Phys. Rev. D* **69**, 123501 (2004).
26. Koushiappas, S. M., Zentner, A. R. & Walker, T. P. Observability of gamma rays from neutralino annihilations in the Milky Way substructure. *Phys. Rev. D* **69**, 043501 (2004).
27. Stoehr, F., White, S. D. M., Springel, V., Tormen, G. & Yoshida, N. Dark matter annihilation in the halo of the Milky Way. *Mon. Not. R. Astron. Soc.* **345**, 1313–1322 (2003).
28. Peirani, S., Mohayaee, R. & de Freitas Pacheco, J. A. Indirect search for dark matter: Prospects for GLAST. *Phys. Rev. D* **70**, 043503 (2004).
29. Pieri, L., Branchini, E. & Hofmann, S. Difficulty of detecting minihalos via γ rays from dark matter annihilation. *Phys. Rev. Lett.* **95**, 211301 (2005).
30. Springel, V., White, S. D. M., Tormen, G. & Kauffmann, G. Populating a cluster of galaxies — I. Results at z=0. *Mon. Not. R. Astron. Soc.* **328**, 726–750 (2001).

LETTERS

Emergence of preformed Cooper pairs from the doped Mott insulating state in Bi$_2$Sr$_2$CaCu$_2$O$_{8+\delta}$

H.-B. Yang[1], J. D. Rameau[1], P. D. Johnson[1], T. Valla[1], A. Tsvelik[1] & G. D. Gu[1]

Superconductors are characterized by an energy gap that represents the energy needed to break the pairs of electrons (Cooper pairs) apart. At temperatures considerably above those associated with superconductivity, the high-transition-temperature copper oxides have an additional 'pseudogap'. It has been unclear whether this represents preformed pairs of electrons that have not achieved the coherence necessary for superconductivity, or whether it reflects some alternative ground state that competes with superconductivity[1]. Paired electrons should display particle–hole symmetry with respect to the Fermi level (the energy of the highest occupied level in the electronic system), but competing states[2–4] need not show such symmetry. Here we report a photoemission study of the underdoped copper oxide Bi$_2$Sr$_2$CaCu$_2$O$_{8+\delta}$ that shows the opening of a symmetric gap only in the anti-nodal region, contrary to the expectation that pairing would take place in the nodal region. It is therefore evident that the pseudogap does reflect the formation of preformed pairs of electrons and that the pairing occurs only in well-defined directions of the underlying lattice.

Angle-resolved photoemission spectroscopy (ARPES; see Supplementary Information) has been used extensively to study the copper oxide superconductors[5–8]. It has been concluded that in the superconducting state, the energy gap associated with the electron pair has d-wave symmetry with maximum gap in the anti-nodal region, and that in the normal state a pseudogap, also in the same region, coexists with 'Fermi arcs'. These arcs extend out from the nodes (points where the superconducting gap has zero value) with a length proportional to T/T^*, T^* being the pseudogap onset temperature[9]. It is generally assumed that the spectral function, the response to the addition or removal of an electron, is particle–hole symmetric around the Fermi surface at low energies, with a peak at the Fermi level on the Fermi arcs and a local minimum at the Fermi level in the gapped regions away from the arcs[10]. In Bardeen–Cooper–Schrieffer theory, the superconducting gap reflects the formation of electron pairs in parallel with the development of long-range phase coherence. The spectral function associated with the paired electrons will consist of two peaks displaying particle–hole symmetry with respect to the Fermi level (see Supplementary Information).

Several techniques have been used in ARPES to obtain a more representative picture of the complete gap. These include 'symmetrization' of the measured photoemission intensity to produce an identical spectral response in both the occupied and the unoccupied states[11]. A second method uses the fact that the measured photoelectron intensity as a function of energy ω and momentum $\mathbf{k}$, $I(\mathbf{k},\omega)$, is given by

$$I(\mathbf{k},\omega) \propto \int A(\mathbf{k},\omega')f(\omega')R(\omega-\omega')\,\mathrm{d}\omega' \qquad (1)$$

with $A(\mathbf{k},\omega)$ the spectral function, $f(\omega)$ the Fermi distribution function and $R(\omega)$ the experimental resolution. The proportionality indicates

the presence of a matrix element dependent on the photon energy. Some level of information can thus be obtained on states thermally occupied above the Fermi level, by dividing the measured intensity by the appropriate temperature-dependent Fermi function. The procedure enhances the response above the chemical potential, but is complicated by the fact that the measured spectral intensity also reflects the experimental resolution. In the past, either the role of the experimental resolution has been ignored or the raw data have been normalized to a Fermi function convolved with a representative resolution function[11–14].
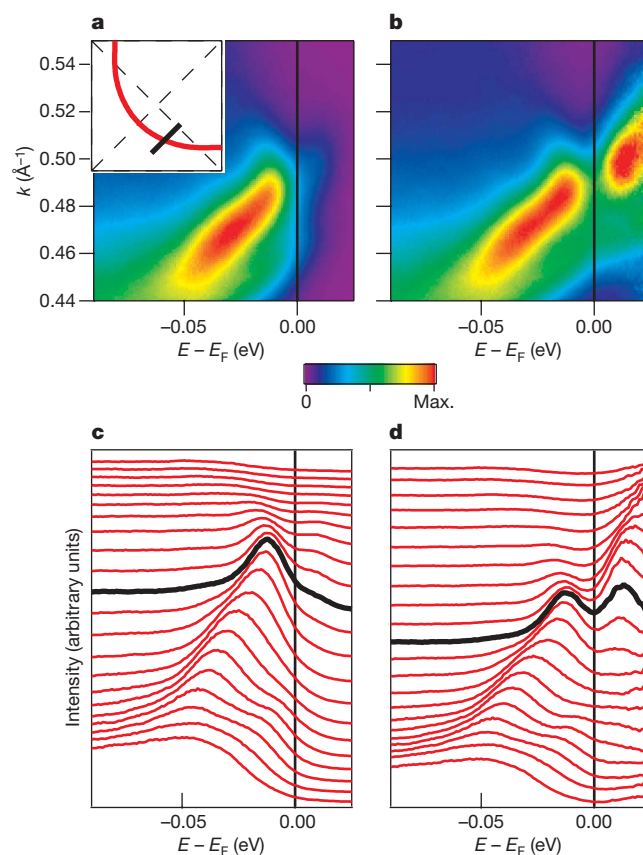

Figure 1 | Analysis of spectra from the optimally doped material. Spectral plots of optimally doped ($T_c = 91$ K) BSCCO as recorded (**a**) and after the analysis described in the text (**b**). The spectra were recorded at a temperature of 80 K and at the point in the Brillouin zone indicated in the schematic in the inset of **a**, where the red line indicates the Fermi surface. The incident photon energy was 16.5 eV. **c**, **d**, EDC cuts through the spectral plots of **a** and **b**, respectively. The EDCs corresponding to k_F are indicated. The spectra in **d** show the dispersion of the Bogoliubov quasi-particles in complete agreement with equation (1).

[1]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA.
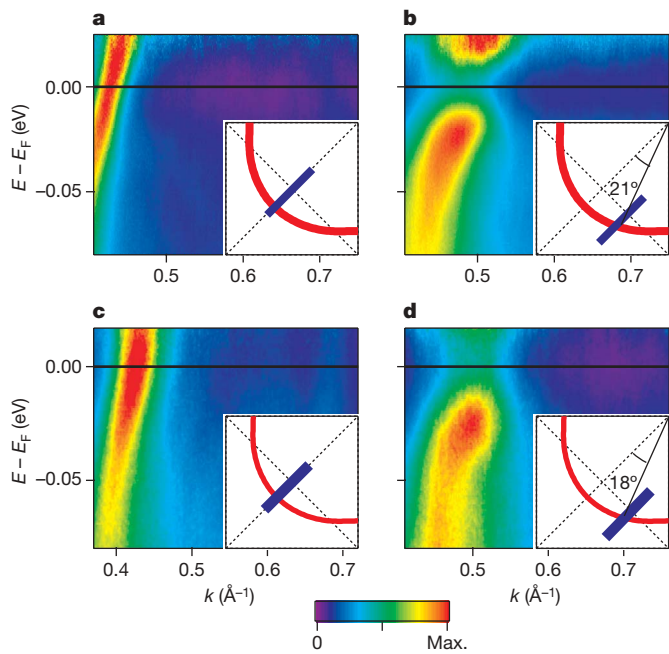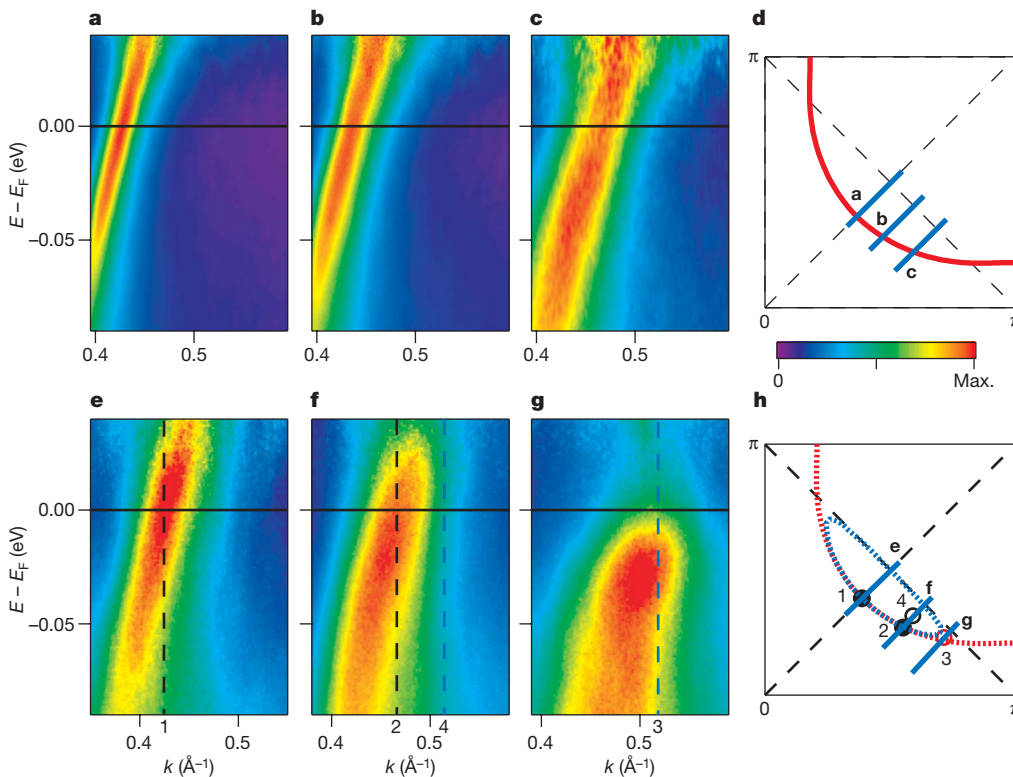
Figure 2 | Spectra from the optimally doped and underdoped material in the superconducting state. Spectral plots after the full analysis as discussed in the text for optimally doped ($T_c = 91$ K) and underdoped ($T_c = 65$ K) BSCCO in the superconducting state. The incident photon energy was 16.5 eV. **a, b,** Spectral plots recorded from the optimally doped sample at a temperature of 80 K in the nodal direction (**a**) and away from the nodal direction (**b**) as indicated in the Brillouin zones shown in the respective insets. **c, d,** Same as **a** and **b**, but for the underdoped material at a temperature of 50 K, again as indicated in the insets. **b** and **d** show the presence of a symmetric superconducting gap.

These methods can lead to serious errors in the relative intensity above and below the Fermi energy, E_F, and to distortions of band dispersions in the vicinity of the Fermi level.

We develop two different approaches to circumvent these problems. The two methods give nearly identical results for the purposes of this study. The first amounts to an approximate solution of equation (1). A function $S(\mathbf{k}, \omega) = A(\mathbf{k}, \omega)f(\omega)$ is determined by convolving $I(\mathbf{k}, \omega)$ with a function that is effectively the inverse transform of a momentum-independent resolution function $R(\omega)$, the latter assumed to be of Gaussian form (see Supplementary Information). Dividing the obtained $S(\mathbf{k}, \omega)$ by the Fermi distribution function, $f(\omega)$, then provides access to the states above the Fermi level.

The second method (see Supplementary Information) uses the fact that the two-dimensional energy-momentum information recorded in modern electron spectrometers is simply an image captured by an array of pixels. In an ideal world with infinite energy and momentum resolution, each pixel i would capture the relevant information, S_i. We use identical labelling in discussing the two methods, to make the discussion more transparent. Thus, S_i is equivalent to the component of the spectrum $S(\mathbf{k}, \omega)$ captured by pixel i. The finite resolution of the system results in the information S_i being distributed across neighbouring pixels with Gaussian widths defined by the experimental resolution. The energy and angular broadening are simultaneously removed using the so-called Lucy–Richardson iterative technique, which is a procedure frequently used in the analysis of medical and astronomical images[15] (see Supplementary Information).

Figure 1 shows the results of such an analysis applied to the spectral intensity measured from an optimally doped ($T_c = 91$ K) $Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCCO) sample in the superconducting state. The dispersion of the Bogoliubov quasi-particles above and below the Fermi level with the transfer of intensity from the occupied to unoccupied states at $\mathbf{k}_F$, the Fermi wavevector, is in complete accord with the Bardeen–Cooper–Schrieffer spectral function (see Supplementary Information).

Figure 2 compares the behaviour in the underdoped and optimally doped systems in the superconducting state. For both samples, the spectra show the opening of a symmetric gap on moving away from the node. Figure 3 shows similar spectra but now in the normal state, above T_c. The spectra from the optimally doped sample show

Figure 3 | Spectra from the optimally doped and underdoped material in the normal state. Spectral plots after the full analysis as discussed in the text for optimally doped ($T_c = 91$ K) and underdoped ($T_c = 65$ K) BSCCO in the normal state. The incident photon energy was 16.5 eV and in all cases the spectra were recorded at a temperature of 140 K. **a, b, c,** Plots recorded from the optimally doped material in the nodal direction and away from the nodal direction, as indicated in **d. e, f, g,** Same as **a, b** and **c**, but for the underdoped material, as indicated in **h**. The magnetic zone boundary would lie at $k = 0.58$ Å^{-1}. In **e** and **f**, the vertical black dashed line indicates the Fermi surface crossing. In **f** and **g**, the vertical blue dashed line indicates the turning point at the top of the dispersion. These are indicated in **h** by the open circles, black indicating a turning point above the Fermi level and red a turning point below the Fermi level. The filled black circles indicate the position of Fermi crossings. The possible pocket is indicated by the area enclosed by the blue dashed line in **h**.
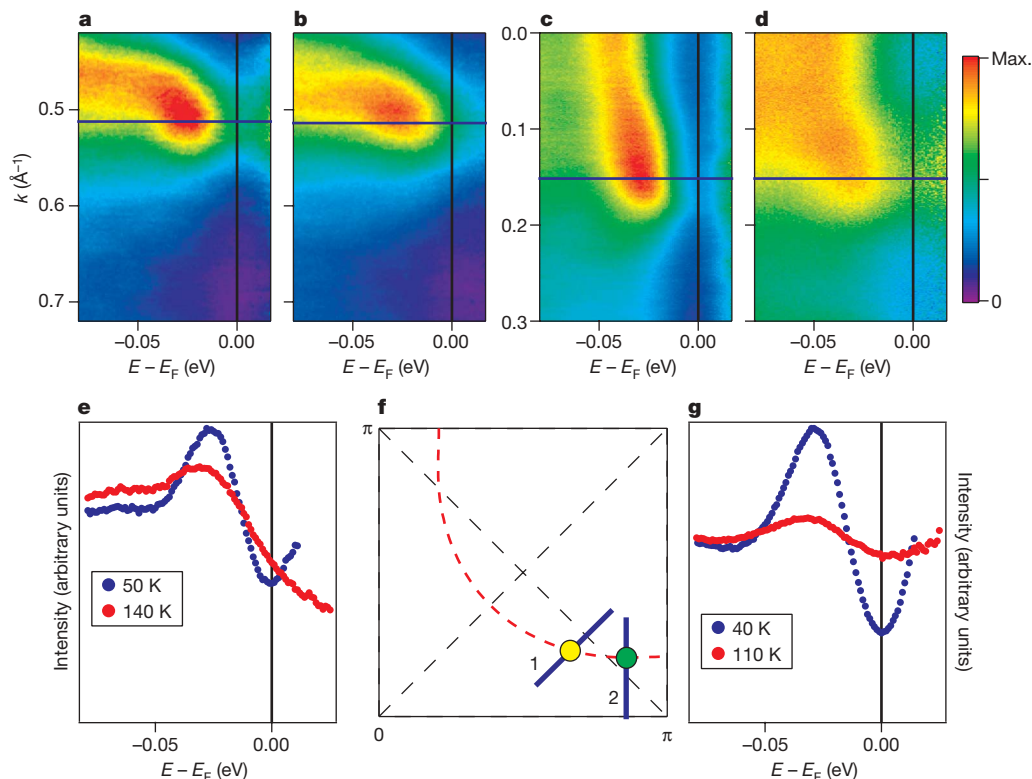
Figure 4 | Analysis of spectra recorded along the anti-nodal direction.
Comparison of the spectral plots after analysis, and associated EDCs for two
different regions of the Brillouin zone, both in the normal and
superconducting states for the underdoped 65 K sample. These are shown in
a, **b** and **e** for the region showing particle–hole asymmetry and in **c**, **d** and
g for the region near the anti-node, as indicated in **f** (points 1 and 2,
respectively). **a**, Spectral intensity recorded from point 1 at a temperature of
50 K; **b**, same as **a**, but in the normal state at a temperature of 140 K.
c, Spectral plot from point 2 recorded at 40 K; **d**, same as **c**, but in the normal
state at 110 K. **e**, EDCs cut through the plots associated with point 1 show
particle–hole symmetry in the superconducting state and asymmetry in the
normal state. **g**, EDCs cut through the plots associated with point 2 show
particle–hole symmetry in both the superconducting and normal states.

the closing of the gap, as might be anticipated. However, in the
normal-state spectra recorded from the underdoped material, the
striking observation is the loss of particle–hole symmetry and the
appearance of a gap above the chemical potential. In particular, it
appears that on moving away from the node a gap appears in the
spectrum and moves down to straddle E_F farther from the node. If
the spectrum were symmetrized in energy at $k = |\mathbf{k}_F|$ in Fig. 3f, it
would incorrectly indicate the presence of a particle–hole symmetric
pseudogap.

The observation of a gap above E_F, together with the absence of
particle–hole symmetry in the underdoped sample, suggests the
absence of pairing in the nodal region in the normal state. There
are several different theoretical explanations for the gap at positive
energies. Some suggest a competing order associated with the under-
lying antiferromagnetism, with the gap reflecting the magnetic zone
boundary[16,17]. Others suggest that the normal state represents a dis-
ordered spin liquid[18–20]. This in turn represents a particle with spin
moving through a sea of spins (the Mott insulating state) rather than
representing a Fermi liquid, where an electron or hole moves through
a sea of electrons. In all of the models, the observed Fermi arc actually
represents the inner half of a hole pocket, the outer half being sup-
pressed by coherence factors, similar to the suppression of the
Bogoliubov quasi-particles observed in Fig. 1. The two sides of the
pocket are defined by bands dispersing up through the Fermi level,
turning back at the gap edge and dispersing down through the Fermi
level again. A representative pocket is indicated by the dashed blue
line in Fig. 3h. The present experimental observations of a folding in
the dispersion (representative energy distribution curves (EDCs) are
shown in the Supplementary Information) followed by a loss of
intensity is fully consistent with such a picture. However, we are able
to make one other important observation based on Fig. 3f, g. As we

discussed earlier, because of the coherence factors we see only one
side of the Fermi pocket. The point at which the dispersing band
ends, seemingly abruptly, not only defines the lower edge of the
gap but also provides an approximate indication of where the other
side of the pocket is located, the two sides of the pocket being
approximately symmetric around the turnover point[18,20]. The fact
that the turnover points are not centred on the magnetic zone bound-
ary effectively rules out models involving broken symmetries reflect-
ing scattering vectors of the type $\mathbf{Q}(\pi, \pi)$; these result in pockets
symmetric around that boundary.

The question then arises as to whether the pairing of electrons
above the superconducting transition temperature, T_c, that has been
observed in recent experiments[21–23] is also evident in ARPES.
Evidence for such phenomena is found in the anti-nodal region
(Fig. 4). There the normalized spectra, for temperatures above and
below T_c, are shown for two different points on the Fermi surface.
The points both show the symmetric gap associated with paired
electrons in the superconducting state but differ markedly in the
normal state, with the spectrum from point 1 showing asymmetric
behaviour and the spectrum from point 2, which is closer to the anti-
nodal region, indicating a symmetric gap. The latter apparent sym-
metry around the Fermi level, similar to that observed in Fig. 1, is a
strong indication of the pairing of electrons along the copper–oxygen
bond directions in the normal state. Such an observation is consistent
with theories that predict the pairing to be essentially one-dimen-
sional in nature[24,25].

1. Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an
    experimental survey. *Rep. Prog. Phys.* **62**, 61–122 (1999).

2. Valla, T. *et al.* Quasiparticle spectra, charge-density waves, superconductivity, and electron-phonon coupling in 2H-NbSe$_2$. *Phys. Rev. Lett.* **92**, 086401 (2004).

3. Schäfer, J. *et al.* Direct spectroscopic observation of the energy gap formation in the spin density wave phase transition at the Cr(110) surface. *Phys. Rev. Lett.* **83**, 2069–2072 (1999).

4. Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).

5. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).

6. Lee, W. S. *et al.* Abrupt onset of a second energy gap at the superconducting transition of underdoped Bi2212. *Nature* **450**, 81–84 (2007).

7. Valla, T., Fedorov, A. V., Lee, J., Davis, J. C. & Gu, G. D. The ground state of the pseudogap in cuprate superconductors. *Science* **314**, 1914–1916 (2006).

8. Tanaka, K. *et al.* Distinct Fermi-momentum-dependent energy gaps in deeply underdoped Bi2212. *Science* **314**, 1910–1913 (2006).

9. Kanigel, A. *et al.* Evolution of the pseudogap from Fermi arcs to the nodal liquid. *Nature Phys.* **2**, 447–451 (2006).

10. Norman, M. R. *et al.* Modeling the Fermi arc in underdoped cuprates. *Phys. Rev. B* **76**, 174501 (2007).

11. Norman, M. R. *et al.* Destruction of the Fermi surface in underdoped high-T_c superconductors. *Nature* **392**, 157–160 (1998).

12. Kisker, E. *et al.* Evidence for the high-spin to low-spin state transition in ordered Fe3Pt Invar. *Phys. Rev. Lett.* **58**, 1784–1787 (1987).

13. Greber, T. *et al.* Photoemission above the Fermi level: the top of the minority *d* band in nickel. *Phys. Rev. Lett.* **79**, 4465–4468 (1997).

14. Matsui, H. *et al.* BCS-like Bogoliubov quasiparticles in high-T_C superconductors observed by angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **90**, 217002 (2003).

15. Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745–754 (1974).

16. Matsui, H. *et al.* Angle-resolved photoemission spectroscopy of the antiferromagnetic superconductor Nd$_{1.87}$Ce$_{0.13}$CuO$_4$: anisotropic spin-correlation gap, pseudogap, and the induced quasiparticle mass enhancement. *Phys. Rev. Lett.* **94**, 047005 (2005).

17. Chakravarty, S. *et al.* Angle-resolved photoemission spectra in the cuprates from the *d*-density wave theory. *Phys. Rev. B* **68**, 100504 (2003).

18. Wen, X. G. & Lee, P. A. Theory of quasiparticles in the underdoped high-T_C superconducting state. *Phys. Rev. Lett.* **80**, 2193–2196 (1998).

19. Konik, R. M., Rice, T. M. & Tsvelik, A. M. Doped spin liquid: Luttinger sum rule and low temperature order. *Phys. Rev. Lett.* **96**, 086407 (2006).

20. Yang, K. Y., Rice, T. M. & Zhang, F. C. Phenomenological theory of the pseudogap state. *Phys. Rev. B* **73**, 174501 (2006).

21. Xu, Z. A. *et al.* Vortex-like excitations and the onset of superconducting phase fluctuation in underdoped La$_{2-x}$Sr$_x$CuO$_4$. *Nature* **406**, 486–488 (2000).

22. Wang, Y. *et al.* Field-enhanced diamagnetism in the pseudogap state of the cuprate Bi$_2$Sr$_2$CaCu$_2$O$_{8+d}$ superconductor in an intense magnetic field. *Phys. Rev. Lett.* **95**, 247002 (2005).

23. Wang, Y., Li, L. & Ong, N. P. Nernst effect in high-T_C superconductors. *Phys. Rev. B* **73**, 024510 (2006).

24. Emery, V. J., Kivelson, S. A. & Zachar, O. Spin-gap proximity effect mechanism of high-temperature superconductivity. *Phys. Rev. B* **56**, 6120–6147 (1997).

25. Tsvelik, A. M. & Chubukov, A. V. Phenomenological theory of the underdoped phase of a high-T_c superconductor. *Phys. Rev. Lett.* **98**, 237001 (2007).

LETTERS

Silicon-chip-based ultrafast optical oscilloscope

Mark A. Foster[1], Reza Salem[1], David F. Geraghty[1], Amy C. Turner-Foster[2], Michal Lipson[2] & Alexander L. Gaeta[1]

With the realization of faster telecommunication data rates and an expanding interest in ultrafast chemical and physical phenomena, it has become important to develop techniques that enable simple measurements of optical waveforms with subpicosecond resolution[1]. State-of-the-art oscilloscopes with high-speed photodetectors provide single-shot waveform measurement with 30-ps resolution. Although multiple-shot sampling techniques can achieve few-picosecond resolution, single-shot measurements are necessary to analyse events that are rapidly varying in time, asynchronous, or may occur only once. Further improvements in single-shot resolution are challenging, owing to microelectronic bandwidth limitations. To overcome these limitations, researchers have looked towards all-optical techniques because of the large processing bandwidths that photonics allow. This has generated an explosion of interest in the integration of photonics on standard electronics platforms, which has spawned the field of silicon photonics[2] and promises to enable the next generation of computer processing units and advances in high-bandwidth communications. For the success of silicon photonics in these areas, on-chip optical signal-processing for optical performance monitoring will prove critical. Beyond next-generation communications, silicon-compatible ultrafast metrology would be of great utility to many fundamental research fields, as evident from the scientific impact that ultrafast measurement techniques continue to make[3–5]. Here, using time-to-frequency conversion[6] via the nonlinear process of four-wave mixing on a silicon chip, we demonstrate a waveform measurement technology within a silicon-photonic platform. We measure optical waveforms with 220-fs resolution over lengths greater than 100 ps, which represent the largest record-length-to-resolution ratio (>450) of any single-shot-capable picosecond waveform measurement technique[6–16]. Our implementation allows for single-shot measurements and uses only highly developed electronic and optical materials of complementary metal-oxide-semiconductor (CMOS)-compatible silicon-on-insulator technology and single-mode optical fibre. The mature silicon-on-insulator platform and the ability to integrate electronics with these CMOS-compatible photonics offer great promise to extend this technology into commonplace bench-top and chip-scale instruments.

Several established nonlinear optical techniques[17,18] exist to measure optical waveforms with few-femtosecond accuracy, but have limited single-shot record lengths of tens of picoseconds and limited update rates. To span the temporal region between electronic measurement and these methods, and to allow for rapidly updateable direct optical detection, techniques have been developed using the space-time duality of electromagnetic waves and related phenomena[6–16]. This duality relies on the equivalence between the paraxial wave equation, which governs diffractive propagation of a spatial field, and the scalar wave equation, which governs dispersive propagation of a temporal field[19,20]. The duality implies that spatial optical components such as a lens or prism have temporal counterparts known as a time-lens or time-prism, which can be implemented by imparting a quadratic or linear temporal phase shift, respectively,

to the temporal field[19,20]. Furthermore, these components allow for temporal processing in a manner analogous to that of the spatial counterparts, such as temporal imaging of the waveform.

Two methods using the space-time duality can be applied to measure ultrafast optical waveforms. Much like a spatial lens can magnify an image, a temporal lens can lengthen an ultrafast waveform in time, allowing for measurement using a photodetector and an oscilloscope that would have insufficient temporal resolution for the unmagnified waveform. This technique is known as temporal magnification[7,8,12]. The second measurement method utilizes the Fourier property of a lens[21]—an object positioned at the front focal plane of a lens will produce a Fourier transform of the object at the back focal plane (Fig. 1a). As the Fourier transform of a temporal waveform is its optical spectrum, extending the spatial Fourier processor to the temporal domain yields a device that converts the temporal (spectral) profile of the input to the spectral (temporal) profile of the output (Fig. 1b). Thus, a measurement of the spectrum at the Fourier plane directly yields the temporal amplitude of the incident waveform, and this process is termed time-to-frequency conversion[6].

The phase shift for temporal imaging devices is typically applied using an electro-optical phase modulator, but an alternative scheme can be realized by using a parametric nonlinear wave-mixing process such as sum-frequency generation and difference-frequency generation. This latter technique is called parametric temporal imaging[22],
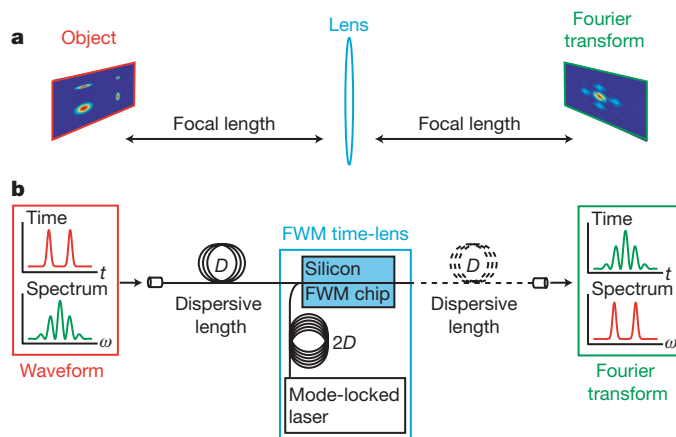

Figure 1 | The silicon-based ultrafast optical oscilloscope. An ultrafast optical oscilloscope is implemented using a four-wave-mixing-based parametric time-lens on a silicon chip. **a**, A spatial optical Fourier transform processor. The spatial lens can generate the Fourier transform of an input waveform using the two-focal-length configuration shown. **b**, A temporal optical Fourier transform processor. The time-lens can convert the temporal profile of the input to the spectral profile of the output. For the FWM time-lens, the focal length (*D*) is half the dispersive length of optical fibre through which the pump pulse passes (2*D*). Single-shot temporal measurements can then be carried out by simply measuring the spectrum at the output of the processor.

[1]School of Applied and Engineering Physics, [2]School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA.

and consists of wave-mixing with a linearly-chirped pump yielding a converted waveform that is nearly equivalent to the signal waveform with a linear frequency chirp or equivalently a quadratic phase shift as required for a time-lens. Parametric time-lenses have phase-shifts in excess of 100π, which is significantly larger than the 10π maximally possible using an electro-optical phase modulator, and therefore greatly extend the applications of temporal imaging systems. A drawback of using the sum-frequency-generation and difference-frequency-generation second-order nonlinear processes is that only a narrow range of materials possess a second-order nonlinear moment, and the converted waveform is inherently generated at widely different wavelengths from that of the pump or input signal. Waveform measurement based on temporal magnification using difference-frequency generation has yielded promising results, including single-shot measurement of ultrafast waveforms with a resolution of less than 900 fs for a simultaneous record length of 100 ps (ref. 12). Waveform measurements based on time-to-frequency conversion using electro-optic modulation have demonstrated a resolution of 3 ps over a 31-ps record length using multiple-shot averaging[6].

Here we demonstrate a parametric time-lens based on the third-order nonlinear process of four-wave mixing (FWM), and apply this time-lens to the creation of a silicon-chip-based ultrafast optical oscilloscope. As our device is based on the third-order Kerr nonlinearity, the FWM-based time-lens can be implemented in any material platform, including the CMOS-compatible silicon-on-insulator (SOI) photonic platform used here. The output of this time-lens is generated at a wavelength close to those of the pump and input waves, enabling all the interacting waves to be in the S, C and L telecommunications bands, for example, which allows for the manipulation of all the waves using the well-established instrumentation and components available for these bands. Using our device, we perform measurements of highly complex waveforms with 220-fs resolution over record lengths larger than 100 ps. The combination of this 220-fs resolution and greater than 100-ps record length represents the largest record-length-to-resolution ratio ($>$450) of any single-shot-capable waveform measurement technique for the picosecond time range[6–16]. Furthermore, unlike commonly used techniques such as frequency-resolved optical gating[17] and spectral-phase interferometry for direct electric-field reconstruction[16,18], our implementation directly measures the temporal amplitude profile using no reconstruction algorithm, allowing for rapidly updateable single-shot measurements.

We test the capability of the silicon-chip-based ultrafast optical oscilloscope with various input waveforms. Each input waveform enters the device and passes through a dispersive element consisting a length of optical fibre. To match to the focal length of the FWM time-lens, the input wave is mixed with a pump pulse that passes through twice the dispersive length of optical fibre. After passing through the optical fibre, the pump pulse and test waveform are combined and FWM is carried out in an SOI nanowaveguide. The strong optical confinement of these silicon structures allows for highly efficient nonlinear processes and for engineerable group-velocity dispersion that can yield conversion bandwidths greater than 150 nm with broad pump tunability[23–27]. The resulting FWM-generated spectrum is measured using an optical spectrometer to determine the temporal profile of the input.

The pump-pulse bandwidth and the length of the dispersive path determine the record length and resolution of the oscilloscope. The time-to-frequency conversion factor for the FWM-based converter is given by

$$\frac{\Delta t}{\Delta \omega} = -\beta_2 L \qquad (1)$$

where Δt is the temporal shift of the input signal, $\Delta \omega$ is the resulting spectral shift, β_2 is the group-velocity dispersion parameter, and L is the length of the dispersive signal path. For our system, this relation yields a 1-nm shift in converted wavelength for a 5.2-ps shift in

temporal position. Using FWM, we can convert a narrow-band signal over twice the pump bandwidth, which yields the approximate record length τ_{record} for the FWM-based oscilloscope

$$\tau_{\text{record}} = 2\beta_2 L\Omega_{\text{pump}} \qquad (2)$$

where Ω_{pump} is the spectral bandwidth of the pump pulse. The resolution of the oscilloscope is predicted by considering the transfer of a temporal delta function through the instrument's system. This impulse response is precisely the temporal resolution $\tau_{\text{resolution}}$ of the instrument and is given by

$$\tau_{\text{resolution}} = \frac{\tau_{\text{pump}}}{\sqrt{2}} \qquad (3)$$

where τ_{pump} is the pump pulse width. For our system, these relationships predict a record length of 150 ps and a resolution of 200 fs. Practically, the separation between the pump and signal and the FWM conversion bandwidth will limit the record length. Deviation from a quadratic phase on the pump pulse, such as that resulting from third-order dispersion, the FWM conversion bandwidth, and the spectral resolution of the spectrometer will also determine the temporal resolution. Since the FWM conversion bandwidth limits both the record length and the resolution, it is important to maximize this value. The silicon waveguides used in our implementation have sufficiently large conversion bandwidths ($>$150 nm)[27] to allow the performance of the ultrafast optical oscilloscope to be solely limited by the aberrations caused by third-order dispersion and the spectrometer performance.

We experimentally characterize the record length and resolution of our system by injecting a 342-fs pulse and varying its temporal position. As shown in Fig. 2, we are able to measure the pulse position across a record length of 100 ps. To characterize the resolution of the FWM-based oscilloscope, we deconvolve the temporal resolution from the average observed width of this pulse across the record length of the device. We measure an average pulse width of 407 fs, which, when compared with the actual pulse width of 342 fs, indicates a temporal resolution of 220 fs for our implementation.

We further investigate the measurement capabilities of the silicon-chip-based oscilloscope by generating test waveforms of varying complexity. First, we measure a pulse which has undergone nonlinear spectral broadening and dispersion using a silicon-chip-based ultrafast optical oscilloscope that exhibits 450-fs resolution and a 100-ps record length. The measurement of this pulse using an ultrafast optical oscilloscope compared with a cross-correlation is shown in Fig. 3a. We measure an optical waveform of even greater complexity by generating a 120-ps waveform with 900-fs temporal features. We measure this waveform using the silicon-chip-based ultrafast optical oscilloscope with 220-fs resolution. The results of this measurement and a comparison to cross-correlation are shown in Fig. 3b.
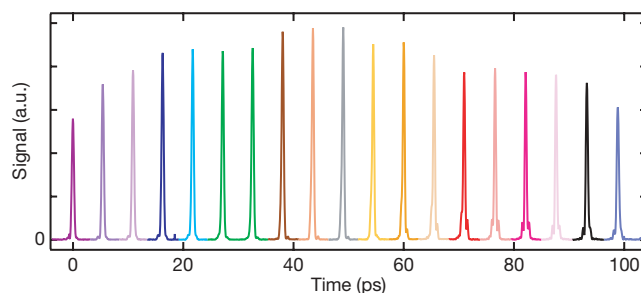
Figure 2 | Characterization of the record length and resolution of the ultrafast oscilloscope. A 342-fs pulse is temporally scanned and measured using the silicon-chip-based ultrafast optical oscilloscope demonstrating a record length of 100 ps. The average width of the 342-fs pulse across this scan range, as observed by the oscilloscope, is 407 fs, indicating a deconvolved resolution of 220 fs. a.u., arbitrary units. Each colour represents a separate measurement as the pulse is scanned.
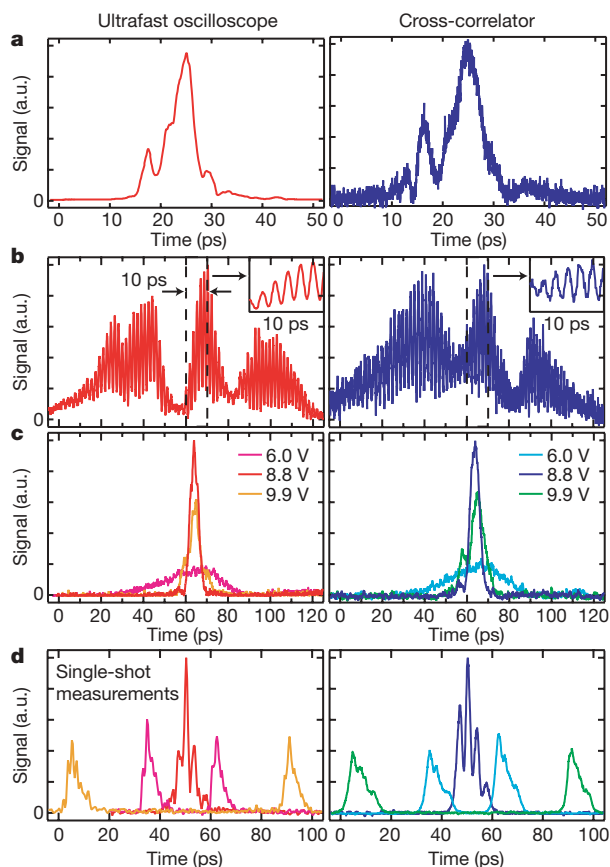
Figure 3 | Comparison of measurements using the ultrafast oscilloscope and a cross-correlator. a, 30-ps pulse generated through nonlinear spectral broadening in an erbium-doped fibre amplifier and subsequently propagated through 20 m of single-mode optical fibre. **b**, Highly complex waveform generated by dispersing and interfering two 300-fs pulses. Inset, magnified view of the 10-ps temporal region from 60 ps to 70 ps. **c**, Measurement of a separate ultrashort-pulse laser source operating at various pulse durations. The silicon-chip-based ultrafast optical oscilloscope is used to minimize the pulse width emitted from this source in real time by varying the voltage to an electro-optic modulator within the laser source. **d**, Single-shot measurements of two chirped pulses with various temporal separations compared with a multiple-shot cross-correlation. When the pulses temporally overlap, interference fringes are observed in the time domain.

The test waveforms in Fig. 3a, b are derived from the same laser source as the pump pulse. We demonstrate that the ultrafast optical oscilloscope can also be used to perform measurements of waveforms from a separate source by synchronizing a variable-pulse-width time-lens-compressed laser source[28] with a repetition rate of 9.6 GHz with the ultrafast fibre laser pump source operating at 38 MHz. Using the device with 220-fs resolution, we optimize the pulse width of the 9.6-GHz source by observing it compress a 30-ps pulse with 30-mW peak power to a 6-ps pulse with 150-mW peak power. The results of this optimization using the ultrafast optical oscilloscope compared to cross-correlation are shown in Fig. 3c.

Lastly, we demonstrate the single-shot capability of the device by incorporating a single-shot spectrometer. We measure three single-shot optical waveforms composed of two pulses with temporal separations of 86 ps, 27 ps, and nearly temporally overlapped. The results of these single-shot measurements compared with a multiple-shot cross-correlation are shown in Fig. 3d. As shown by the 86-ps separation, we maintain the 100-ps record length. When the pulses overlap, we observe temporal interference fringes with a 3-ps period. For this implementation, the temporal resolution is limited to 766 fs per pixel, or a record-length-to-resolution ratio of 130, by the infrared camera. High dynamic range linear arrays with more than 1,000

pixels are commercially available, and would allow for utilization of the full (>450) record-length-to-resolution ratio of our device.

Ultimately, the dynamic range for single-shot measurements is limited by the maximum power allowed in the silicon nanowaveguide while avoiding self-phase modulation and free-carrier generation[27], and by the minimum detectable power per pixel. These constraints should limit the range of signal peak power at the time-lens from 100 µW to 100 mW, which corresponds to a dynamic range of 10^3. The maximum power into the ultrafast optical oscilloscope is dependent on the feature width, as a narrow temporal feature will spread during dispersive propagation before the lens and therefore the peak power at the lens is significantly lower. If resolution-limited temporal features are considered, a 40 W peak power is allowed, which corresponds to a dynamic range of 10^5. Furthermore, because the minimum detectable power depends on the desired single-shot resolution while the maximum power does not, higher dynamic range measurements are possible in this system at reduced resolution.

In all of our measurements, we observe good agreement between our silicon-based ultrafast optical oscilloscope and the cross-correlation with a 280-fs pulse. Nevertheless some deviations are observed, which partially result from the slightly different lengths (less than 3-m variation) of optical fibre used to synchronize the arrival time of the waveforms and pump pulses to the cross-correlator as compared with the oscilloscope. Further inconsistencies are probably due to pump pulse imperfections in the FWM time-lens. For optimal performance, care must be taken to obtain a clean and flat spectral amplitude and phase for the pump pulse. Moreover, the resolution is ultimately limited by the aberrations arising from third-order dispersion in the dispersive elements. The use of dispersion-flattened fibre or dispersion-engineered waveguides[23,24] in the dispersive paths would alleviate this aberration, and provide a path towards sub-100-fs resolution by using a sub-100-fs pump pulse.

The components of this measurement system can potentially be entirely integrated on-chip. Specifically, the integration of a pulsed laser source[29], low-loss dispersion engineered waveguides for the dispersive paths[23,24], and an integrated single-shot spectrometer and detectors[30] are all areas of current research in silicon photonics. Furthermore, the flexibility of the FWM time-lens and the dispersion engineering available in nanowaveguides allow for straightforward extension of this technique to different wavelength regimes (for example the visible) by using other CMOS-compatible waveguiding materials such as SiN and SiON. Additionally, using our oscilloscope for measuring an arbitrary-repetition-rate source requires an ultrafast pump laser with repetition-rate flexibility and can be implemented, for example, using a time-lens compressed source[28]. Interestingly, the single-shot capability will not only allow for measurements of single optical events but, when synchronized with an optical clock, will also allow for measurements of 'eye-diagrams' by overlaying many single-shot measurements of a communications signal. Beyond communications, an integrated measurement device would facilitate studies in many branches of science where simple, ultrafast measurements of optical waveforms are required.

METHODS SUMMARY

To experimentally characterize the silicon-based ultrafast optical oscilloscope, we generate the pump and input waves from an ultrafast fibre laser or an optical parametric oscillator. The pulse train is spectrally separated into a 280-fs pump pulse and a signal pulse. Each input waveform enters the oscilloscope and passes through a dispersive element consisting of a 50-m length of dispersion compensation fibre and is mixed with a pump pulse that has been passed through a 100-m length of dispersion compensation fibre. The test waveforms in Fig. 3a–c were created using combinations of nonlinear spectral broadening, dispersion, and interference. The 1.5-cm-long silicon nanowaveguide has a cross-sectional size of 300 nm by 750 nm, a linear propagation loss of 1.5 dB cm^{-1}, and a 3-dB coupling efficiency. For multiple-shot measurements, the FWM optical spectrum is characterized using an optical spectrum analyser. For the single-shot demonstration, a single-shot spectrometer is implemented using a monochromator and infrared camera and a single event is created per frame.

1. Dorrer, C. High-speed measurements for optical telecommunication systems. *IEEE Select. Topics Quant. Electron.* **12,** 843–858 (2006).
2. Jalali, B. Can silicon change photonics? *Phys. Status Solidi* **205,** 213–224 (2008).
3. Dudley, J. M., Finot, C., Richardson, D. J. & Millot, G. Self-similarity in ultrafast nonlinear optics. *Nature Phys.* **3,** 597–603 (2007).
4. Solli, D. R., Ropers, C., Koonath, P. & Jalali, B. Optical rogue waves. *Nature* **450,** 1054–1057 (2007).
5. Solli, D. R., Chou, J. & Jalali, B. Amplified wavelength-time transformation for real-time spectroscopy. *Nature Photon.* **2,** 48–51 (2008).
6. Kauffman, M. T., Banyal, W. C., Godil, A. A. & Bloom, D. M. Time-to-frequency converter for measuring picosecond optical pulses. *Appl. Phys. Lett.* **64,** 270–272 (1994).
7. Bennett, C. V., Scott, R. P. & Kolner, B. H. Temporal magnification and reversal of 100 Gb/s optical data with an upconversion time microscope. *Appl. Phys. Lett.* **65,** 2513–2515 (1994).
8. Bennett, C. V. & Kolner, B. H. Upconversion time microscope demonstrating 103× magnification of femtosecond waveforms. *Opt. Lett.* **24,** 783–785 (1999).
9. Mouradian, L. K., Louradour, F., Messager, V., Barthelemy, A. & Froehly, C. Spectro-temporal imaging of femtosecond events. *IEEE J. Quant. Electron.* **36,** 795–801 (2000).
10. Azana, J., Berger, N. K., Levit, B. & Fischer, B. Spectral Fraunhofer regime: Time-to-frequency conversion by the action of a single time lens on an optical pulse. *Appl. Opt.* **43,** 483–490 (2004).
11. Fernandez-Pousa, C. R. Temporal resolution limits of time-to-frequency transformations. *Opt. Lett.* **31,** 3049–3051 (2006).
12. Bennett, C. V., Moran, B. D., Langrock, C., Fejer, M. M. & Ibsen, M. 640 GHz real time recording using temporal imaging. In *Conference on Lasers and Electro-Optics* [CD] paper CtuA6 (OSA Technical Digest Series, Optical Society of America, 2008).
13. Kan'an, A. M. & Weiner, A. M. Efficient time-to-space conversion of femtosecond optical pulses. *J. Opt. Soc. Am. B* **15,** 1242–1245 (1998).
14. Oba, K., Sun, P. C., Mazurenko, Y. T. & Fainman, Y. Femtosecond single-shot correlation system: A time-domain approach. *Appl. Opt.* **38,** 3810–3817 (1999).
15. Chou, J., Boyraz, O. & Jalali, B. Femtosecond real-time single-shot digitizer. *Appl. Phys. Lett.* **91,** 161105 (2007).
16. Bromage, J., Dorrer, C., Begishev, I. A., Usechak, N. G. & Zuegel, J. D. Highly sensitive, single-shot characterization for pulse widths from 0.4 to 85 ps using electro-optic shearing interferometry. *Opt. Lett.* **31,** 3523–3525 (2006).
17. Kane, D. J. & Trebino, R. Single-shot measurement of the intensity and phase of an arbitrary ultrashort pulse by using frequency-resolved optical gating. *Opt. Lett.* **18,** 823–825 (1993).
18. Dorrer, C. *et al.* Single-shot real-time characterization of chirped-pulse amplification systems by spectral phase interferometry for direct electric-field reconstruction. *Opt. Lett.* **24,** 1644–1646 (1999).
19. Akhmanov, S. A., Vysloukh, V. A. & Chirkin, A. S. Self-action of wave packets in a nonlinear medium and femtosecond laser pulse generation. *Sov. Phys. Usp.* **29,** 642–677 (1986).
20. Kolner, B. H. Space-time duality and the theory of temporal imaging. *IEEE J. Quant. Electron.* **30,** 1951–1963 (1994).
21. Goodman, J. W. *Introduction to Fourier Optics* (McGraw-Hill, 1968).
22. Bennett, C. V. & Kolner, B. H. Principles of parametric temporal imaging—Part I: System configurations. *IEEE J. Quant. Electron.* **36,** 430–437 (2000).
23. Dulkeith, E., Xia, F., Schares, L., Green, W. M. J. & Vlasov, Y. A. Group index and group velocity dispersion in silicon-on-insulator photonic wires. *Opt. Express* **14,** 3853–3863 (2006).
24. Turner, A. C. *et al.* Tailored anomalous group-velocity dispersion in silicon channel waveguides. *Opt. Express* **14,** 4357–4362 (2006).
25. Foster, M. A. *et al.* Broad-band optical parametric gain on a silicon photonic chip. *Nature* **441,** 960–963 (2006).
26. Lin, Q., Zhang, J., Fauchet, P. M. & Agrawal, G. P. Ultrabroadband parametric generation and wavelength conversion in silicon waveguides. *Opt. Express* **14,** 4786–4799 (2006).
27. Foster, M. A., Turner, A. C., Salem, R., Lipson, M. & Gaeta, A. L. Broad-band continuous-wave parametric wavelength conversion in silicon nanowaveguides. *Opt. Express* **15,** 12949–12958 (2007).
28. van Howe, J., Lee, J. H. & Xu, C. Generation of 3.5 nJ femtosecond pulses from a continuous-wave laser without mode locking. *Opt. Lett.* **32,** 1408–1410 (2007).
29. Koch, B. R., Fang, A. W., Cohen, O. & Bowers, J. E. Mode-locked silicon evanescent lasers. *Opt. Express* **15,** 11225–11233 (2007).
30. Cheben, P. *et al.* A high-resolution silicon-on-insulator arrayed waveguide grating microspectrometer with submicrometer aperture waveguides. *Opt. Express* **15,** 2299–2306 (2007).

METHODS

Laser sources. The ultrafast fibre laser used produces 80-fs pulses at a 38-MHz repetition rate. The optical parametric oscillator used produces 150-fs pulses at a 76-MHz repetition rate. The pump pulse is a 280-fs pulse with 15 nm of bandwidth centred at 1,550 nm. The test waveforms for Fig. 3 are generated from a variable bandwidth signal pulse centred at 1,580 nm.

Optical fibre. We chose to use dispersion compensation fibre (Corning model: DCM-D-080-04) as it has a dispersion slope that is 12× smaller than that of standard single-mode fibre (Corning model: SMF-28). This smaller third-order dispersion reduces lens aberrations, and experimentally we find a 2× improvement in the temporal resolution as compared to an equivalent system using SMF-28. After passing through the dispersion compensation fibre, the 15-nm-bandwidth pump pulse is amplified using an erbium-doped fibre amplifier, and subsequently FWM is carried out in a CMOS-compatible embedded SOI nanowaveguide.

Test waveforms. The test waveform in Fig. 3a is created by amplifying the signal pulse in an erbium-doped fibre amplifier and inducing nonlinear spectral broadening in the amplifier. The spectrally broadened pulse is subsequently passed through a 20-m length of optical fibre. The test waveform in Fig. 3b is generated by dispersing and interfering two 300-fs pulses using 50-m of optical fibre and a Michelson interferometer. The test waveforms in Fig. 3c are generated by synchronizing a time-lens compressed laser source[28] with a repetition rate of 9.6 GHz with an ultrafast fibre laser pump source operating at 38 MHz. The pulse width of the 9.6-GHz source is determined by the magnitude of the electrical sine wave sent into a phase modulator used for the time-lens compressor. The test waveforms in Fig. 3d are generated by chirping a 300-fs pulse using 50-m of SMF-28 and splitting it into two pulses using a Michelson interferometer. The separation between the pulses can then be adjusted using a delay stage on the interferometer.

Silicon waveguide. The dimensions of the silicon waveguide were chosen to maximize the conversion bandwidth by positioning a zero-group-velocity dispersion point in the C telecommunications band. The peak optical power inside the nanowaveguides is maintained below 100 mW to avoid self-phase modulation and two-photon induced free-carrier effects in the silicon[27].

Single-shot measurements. A single event is created per frame of the single-shot spectrometer. The 38-MHz source is down-sampled using an electro-optical modulator such that only one pulse is generated every 0.5 µs, which corresponds to the integration time of the camera and therefore a single shot per camera image.

LETTERS

Atlantic overturning responses to Late Pleistocene climate forcings

Lorraine E. Lisiecki[1]†, Maureen E. Raymo[1] & William B. Curry[2]

The factors driving glacial changes in ocean overturning circulation are not well understood. On the basis of a comparison of 20 climate variables over the past four glacial cycles, the SPECMAP project[1] proposed that summer insolation at high northern latitudes (that is, Milankovitch forcing) drives the same sequence of ocean circulation and other climate responses over 100-kyr eccentricity cycles, 41-kyr obliquity cycles and 23-kyr precession cycles. SPECMAP analysed the circulation response at only a few sites in the Atlantic Ocean, however, and the phase of circulation response has been shown to vary by site and orbital band[2]. Here we test the SPECMAP hypothesis by measuring the phase of orbital responses in benthic $\delta^{13}C$ (a proxy indicator of ocean nutrient content) at 24 sites throughout the Atlantic over the past 425 kyr. On the basis of $\delta^{13}C$ responses at 3,000–4,010 m water depth, we find that maxima in Milankovitch forcing are associated with greater mid-depth overturning in the obliquity band but less overturning in the precession band. This suggests that Atlantic overturning is strongly sensitive to factors beyond ice volume and summer insolation at high northern latitudes. A better understanding of these processes could lead to improvements in model estimates of overturning rates, which range from a 40 per cent increase to a 40 per cent decrease at the Last Glacial Maximum[3] and a 10–50 per cent decrease over the next 140 yr in response to projected increases in atmospheric CO_2 (ref. 4).

Different modes of Atlantic overturning appear to be coupled with widespread climate change on both orbital[1,5,6] and millennial[7,8] time-scales. Today, North Atlantic Deep Water (NADW) fills most of the Atlantic at 2,000–4,000 m water depth, above deep water of Southern Ocean origin. During the Last Glacial Maximum (LGM), nutrient-poor NADW shoaled to less than 2,000 m and was replaced by nutrient-rich Southern Ocean Water (SOW), according to Atlantic transects of the $\delta^{13}C$ and Cd/Ca ratio of LGM benthic foraminifera[9,10]. Although these two proxies can be influenced by several biogeochemical factors unrelated to ocean circulation[9–11], independent estimates of overturning since the LGM are consistent with these tracer changes[7,12]. Comparisons of Pleistocene Atlantic $\delta^{13}C$ records with the ice-volume proxy benthic $\delta^{18}O$ show that the presence of NADW below 2,000 m is strongly correlated with the size of northern ice sheets[1,2,13]. However, important differences between $\delta^{13}C$ and $\delta^{18}O$ suggest that ice volume is not the only factor controlling NADW circulation[2,13].

In the sequence of circulation responses described by SPECMAP[1], lower NADW formation, the 'Nordic heat pump', responds rapidly to Milankovitch forcing, and mid-depth (~3,400 m) overturning in the North Atlantic responds later, slightly lagging ice volume. (This mid-depth response is not equivalent to the 'boreal heat pump', which is the mechanism SPECMAP proposed for the production of Glacial North Atlantic Intermediate Water (GNAIW)[6].) We analyse benthic $\delta^{13}C$ records of the past 425 kyr from 24 Atlantic sites and 5

Pacific sites (Supplementary Table 1 and Supplementary Fig. 1) to evaluate the SPECMAP hypothesis that circulation response has the same phase relative to ice volume in all three orbital bands. We focus on mid-depth sites because circulation responses at these sites are more easily distinguished from changes in the $\delta^{13}C$ composition of NADW and SOW.

We place all benthic $\delta^{13}C$ records on the same age model[14] (Methods) and calculate $\Delta\delta^{13}C$ by taking the difference between each Atlantic $\delta^{13}C$ record and a record of mean ocean $\delta^{13}C$, estimated by averaging five Pacific $\delta^{13}C$ records (Fig. 1). Figure 2 shows the phase of $\Delta\delta^{13}C$ at each site relative to obliquity and precession (June perihelion). (See Supplementary Information for eccentricity phases.) Also shown is the phase of benthic $\delta^{18}O$ (times −1), which we refer to as minimum ice volume for simplicity but which also contains a significant deep water temperature component (Supplementary Information).

The orbital phases of Atlantic $\Delta\delta^{13}C$ can be well described by dividing the sites into three groups according to water depth. On the basis of modern and LGM $\delta^{13}C$ transects[9], we interpret the $\Delta\delta^{13}C$ of sites at 1,100–2,301 m depth as primarily recording the $\delta^{13}C$ composition of NADW, sites from 3,000 to 4,010 m as primarily recording changes in the mixing ratio of NADW and SOW, and sites from 4,035 to 4,620 m as primarily recording the $\delta^{13}C$ of SOW. The consistency of phase relationships within these depth intervals, despite a wide range of latitude and longitude, strongly suggests that these responses are basin-wide and that the $\delta^{13}C$ of deep water is accurately recorded.

At all three frequencies, most sites above 2,300 m have large phase lags (140°–260°), consistent with increases in GNAIW and the $\Delta\delta^{13}C$ of upper NADW during glacial conditions[6,9]. However, at most sites from 3,000 to 4,010 m, $\Delta\delta^{13}C$ leads minimum ice volume in the obliquity band but lags it in the precession band. Mid-depth lags relative to Milankovitch forcing, which we define to be 21 June insolation at 65° N (Supplementary Information), are 3°–26° (0–3 kyr) in the obliquity band and 100°–176° (6–11 kyr) in the precession band. This represents a significant challenge to the SPECMAP hypothesis that circulation responds with the same phase in all orbital bands[1]. Evidence that benthic $\delta^{18}O$ change can differ from ice volume change by 2.2 kyr (ref. 15) cannot account for the ~6-kyr shift in $\delta^{13}C$ phase between the obliquity and precession bands.

The difference between obliquity and precession responses was not observed by SPECMAP because they analysed mid-depth $\delta^{13}C$ only at ODP site 607 (ref. 16), which has atypical obliquity and precession phases in comparison with most other mid-depth sites. However, our results are consistent with previously reported phases at several western equatorial Atlantic sites[2]. Mid-depth sites in the Atlantic western boundary current (squares in Fig. 2) tend to have strong precession and obliquity power and large precession lags. Eastern Atlantic sites

[1]Department of Earth Sciences, Boston University, 675 Commonwealth Avenue, Boston, Massachusetts 02215, USA. [2]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. †Present address: Department of Earth Science, University of California, Santa Barbara, Santa Barbara, California 93106-9630, USA.
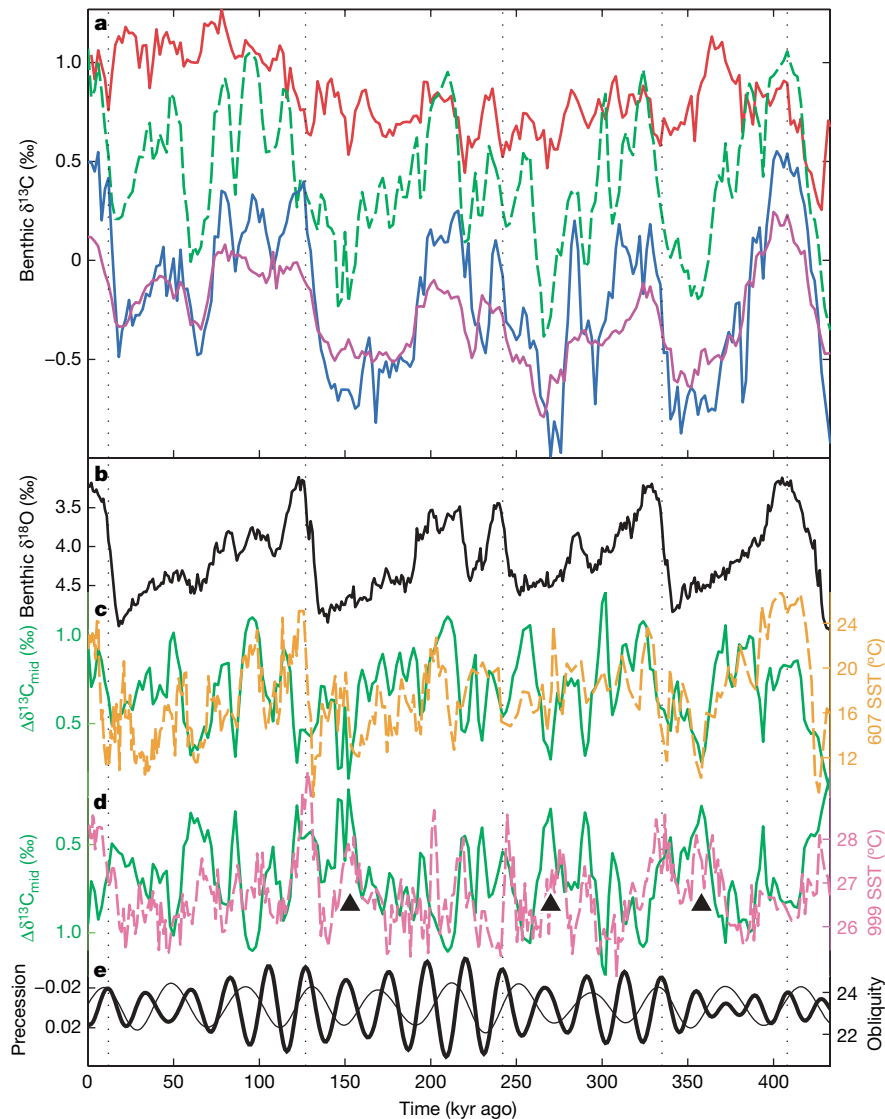
85

Figure 1 | Comparison of benthic δ¹³C, orbital forcing, ice volume and SST. **a**, Regional stacks (averages) of benthic δ¹³C (‰ VPDB) from shallow North Atlantic sites (red; mean from sites DSDP552, ODP980, ODP982, ODP983), selected mid-depth Atlantic sites (green dashed; ODP925, ODP927, ODP928, GeoB1214), deep Atlantic sites (blue; ODP929, ODP1089, GeoB1041, GeoB1211), and Pacific sites (purple; ODP677, ODP846, ODP849, RC13-110, V21-146). **b**, Benthic δ¹⁸O (‰ VPDB; black, ref. 14) as a proxy for ice volume. **c**, North Atlantic summer SST[19] ('607 SST', orange) and $\Delta\delta^{13}C_{mid}$ (green, $\delta^{13}C_{mid} - \delta^{13}C_{Pac}$). **d**, Caribbean SST[21] ('999 SST', pink) and $\Delta\delta^{13}C_{mid}$ (green, *y*-axis reversed). Triangles mark SST peaks preceding glacial maxima. **e**, Precession index (thick) and obliquity (thin), plotted such that greater Milankovitch forcing is up.

with less precession power and smaller precession lags could be less sensitive to circulation change because they are farther from the main path for meridional deep water transport.

To investigate the differences between precession and obliquity responses, we calculate $\Delta\delta^{13}C_{mid}$ by averaging the four mid-depth sites with the greatest precession lags (Fig. 1c). $\Delta\delta^{13}C_{mid}$ displays significantly different phases with respect to ice volume in the obliquity and precession bands (Fig. 3). In the obliquity band, $\Delta\delta^{13}C_{mid}$ is nearly in phase with Milankovitch forcing, suggesting a rapid response to obliquity-driven insolation change. In the precession band, $\Delta\delta^{13}C_{mid}$ is not easily interpreted as a response to Milankovitch forcing or ice volume because it lags June perihelion by 170° (11 kyr) and minimum ice volume by 100° (6.4 kyr). Because circulation is unlikely to lag its forcing by more than 90°, we discuss the alternative possibility that the insolation forcing at June perihelion produces a minimum in $\Delta\delta^{13}C_{mid}$ with very little lag.

Is the minimum in $\Delta\delta^{13}C_{mid}$ at June perihelion caused by a change in circulation? Mid-depth δ¹³C is affected by the mixing ratio of NADW and SOW and by the δ¹³C composition of each water mass. However, the δ¹³C of NADW as recorded by shallow North Atlantic

sites has much less precession power than $\Delta\delta^{13}C_{mid}$ (Supplementary Fig. 2), and the δ¹³C of SOW as recorded by deep South Atlantic sites has a significantly different phase from $\Delta\delta^{13}C_{mid}$ (Fig. 3b). Therefore, the minimum in $\Delta\delta^{13}C_{mid}$ at June perihelion must primarily reflect a reduction in the mixing ratio of NADW at these sites, probably due to weaker and/or shallower North Atlantic overturning.

Although δ¹³C gradients cannot unequivocally constrain the rate of meridional overturning[17], changes in meridional heat transport, inferred from Atlantic sea surface temperature (SST) records, may provide indirect evidence for changes in overturning rate[8]. For example, during stadial events of the last glacial cycle, reduced overturning is associated with cooling at high northern latitudes[7] and warming at low latitudes[8,18], consistent with a decrease in meridional heat transport. If $\Delta\delta^{13}C_{mid}$ records orbitally driven changes in overturning rate, we expect $\Delta\delta^{13}C_{mid}$ to be in phase with North Atlantic SST and antiphased with low-latitude SST (Supplementary Information).

In the precession band, we find that SST estimates based on foraminiferal species counts at ODP Site 607 (Fig. 1c, ref. 19) and other North Atlantic sites[5,20] are nearly in phase with $\Delta\delta^{13}C_{mid}$ (Fig. 3d),
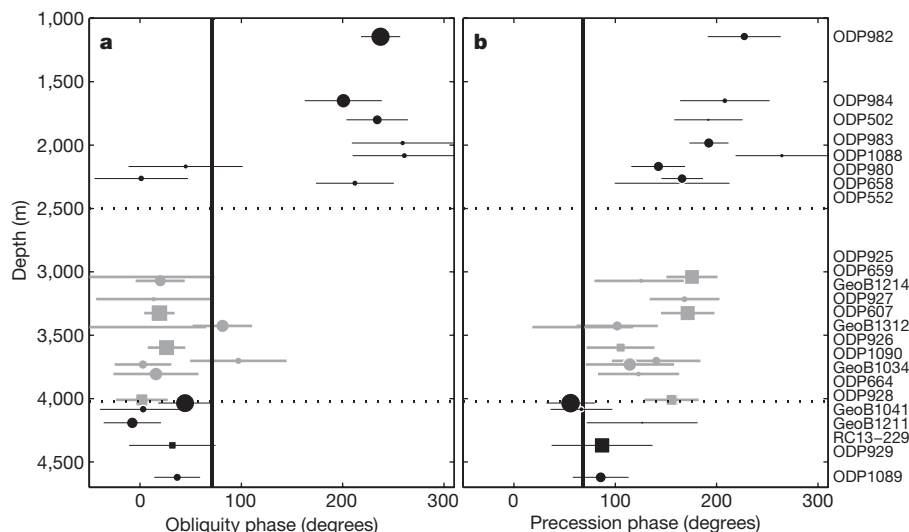
Figure 2 | Phases of orbital responses of benthic $\Delta\delta^{13}C$ in the Atlantic. Symbols mark the phase of each site, with 2σ error bars, relative to **a**, 41-kyr obliquity and **b**, 23-kyr precession (June perihelion). The vertical line in each panel marks the phase of minimum ice volume from benthic $\delta^{18}O$ (ref. 14). Squares denote western Atlantic Ceara rise sites; circles, all other sites. The size of each symbol is proportional to the coherent amplitude of the response. Dotted lines separate different depth ranges; grey symbols denote mid-depth sites. (Site 502 at a depth of 3,051 m in the Caribbean basin is plotted at the sill depth of 1,800 m.)

consistent with a change in overturning rate and out-of-phase with local summer insolation. A Caribbean foraminiferal Mg/Ca record of SST[21] (Fig. 1d) is antiphased with $\Delta\delta^{13}C_{mid}$ (Fig. 3d), consistent with the proposed overturning changes, but also in phase with local summer insolation forcing[21]. We suggest that the effects of overturning can be distinguished from insolation forcing during prominent SST peaks preceding glacial maxima (marked by triangles in Fig. 1d). This low-latitude warming is consistent with the reduced overturning suggested

by $\Delta\delta^{13}C_{mid}$ but difficult to explain otherwise, given relatively weak precession forcing, high ice volume and low atmospheric partial pressure of CO_2 (p_{CO_2}; ref. 22). Therefore, Atlantic SST responses appear consistent with reduced overturning rates during June perihelion, as suggested by $\Delta\delta^{13}C_{mid}$.

In the obliquity band, the phases of Atlantic SST at both high and low latitudes are similar (Fig. 3c) rather than antiphased and, thus, provide no constraints on the phase of overturning response. The fact that both SST records are approximately in phase with ice volume and p_{CO_2} (Supplementary Table 2) suggests that these factors may overwhelm any SST signal due to obliquity-driven overturning changes. In contrast, ice volume and p_{CO_2} (ref. 22) have less power in the precession band, which may explain why the effects of over-turning on SST are visible for precession but not obliquity.

Can modelling studies be used to provide another test of circulation responses to obliquity and precession? It seems not; at present, the response of North Atlantic overturning to orbital forcing or even LGM boundary conditions[3] is highly model-dependent. Increasing Milankovitch forcing decreases NADW formation in some models[23,24] and increases it in others[25,26]. One study that varied obliquity and precession separately found that greater Milankovitch forcing from either precession or obliquity increased overturning[26]. However, another found that greater Milankovitch forcing decreased overturning in the obliquity band but increased it in the precession band[27]. Additionally, neither of these models included a Laurentide ice sheet. In fact, given the wide range of model results, our results may provide an important metric against which future models could be evaluated; our results could be further tested with additional SST records and other circulation proxies.

The phases of $\Delta\delta^{13}C_{mid}$ place constraints on the forcing mechanisms important for mid-depth overturning. Greater Milankovitch forcing decreases ice volume in both the obliquity and precession bands but corresponds to different $\Delta\delta^{13}C_{mid}$ responses. Therefore, mid-depth overturning must also be sensitive to factors other than Milankovitch forcing and ice volume. The different seasonal and spatial insolation anomalies associated with precession versus obliquity provide many mechanisms by which the two orbital cycles could produce different overturning responses.

The strongest insolation forcing that is antiphased with Milankovitch forcing in one orbital band but not the other occurs at high southern latitudes in summer. Weaker forcing occurs over
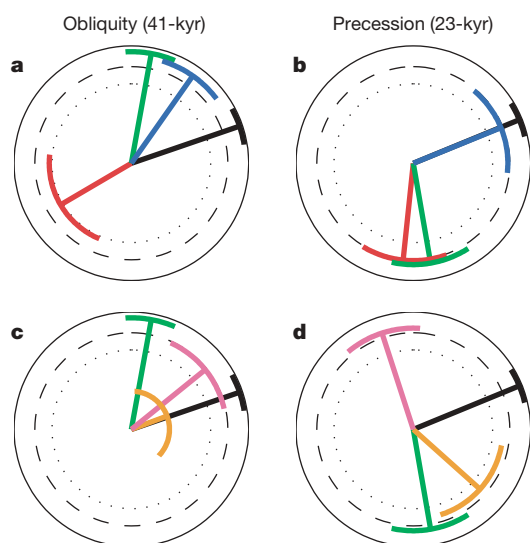

Figure 3 | Comparison of $\Delta\delta^{13}C$ and SST phases. a, b, The phases for obliquity (**a**) and precession (**b**) of minimum ice volume from benthic $\delta^{18}O$ (black, ref. 14) and regional stacks of Atlantic $\Delta\delta^{13}C$ for shallow North Atlantic sites (red), selected mid-depth sites (green) and >4,010 m sites (blue). **c, d**, The phases for obliquity (**c**) and precession (**d**) of benthic $\delta^{18}O$ (black, ref. 14), $\Delta\delta^{13}C_{mid}$ (green), North Atlantic SST[19] (orange) and Caribbean SST[21] (pink). In this phase wheel representation, vectors in the 12 o'clock position are in phase with maximum Milankovitch forcing, and phase lags increase in the clockwise direction (for example, 3 o'clock represents a 90° lag relative to Milankovitch forcing, 6 o'clock represents an antiphased response, and 9 o'clock represents a 90° lead). Vector length (from circle centre to middle of arc) represents coherence, and the associated arc denotes the 2σ phase error. Circles mark 100% (solid), 95% (dashed) and 80% (dotted) coherence.

southern mid-latitudes in summer, southern low- and mid-latitudes in winter, and high northern latitudes in late summer (Supplementary Fig. 7). Perhaps cooler southern summers during June perihelion, obliquity minima, and glacial maxima reduce mid-depth NADW by enhancing SOW formation[28]. However, Antarctic temperature appears to be in phase with Milankovitch forcing rather than local summer insolation[29]. Alternatively, reduced insolation in late summer at high northern latitudes[30] might decrease NADW formation by altering North Atlantic sea ice extent, salinity or temperature. The different phases of circulation response relative to Milaukovitch forcing associated with obliquity and precession may produce different climate feedbacks, which may affect the amplitude of ice volume response at the two frequencies.

METHODS SUMMARY

The $\delta^{13}C$ records in this study are collected from previous studies (Supplementary Information) and primarily measured from the epibenthic taxon *Cibicidoides wuellerstorfi*. All records are placed on a common age model by aligning their benthic $\delta^{18}O$ records to the LR04 benthic $\delta^{18}O$ stack[14]. (Site ODP999 (ref. 21) was aligned using planktonic $\delta^{18}O$.) See ref. 14 for detailed alignment methodology.

We reconstruct circulation changes at mid-depth sites using the $\Delta\delta^{13}C$ instead of percentage NADW[1] because we wish to avoid the assumption that the $\delta^{13}C$ of SOW and NADW are consistently recorded by any of the available sites. In particular, many of our deep Atlantic (SOW) sites have a non-negligible percentage NADW today. However, the phases of percentage NADW for mid-depth sites (Supplementary Fig. 3) are not significantly different from those of $\Delta\delta^{13}C$. Additionally, the phases of $\Delta\delta^{13}C$ are not significantly different from $\delta^{13}C$ (Supplementary Fig. 4), except at several shallow Atlantic sites.

Before spectral analysis, all records are interpolated to an even 1-kyr time step from 15 to 425 kyr ago. Spectral analysis was performed with the ARAND software package (P. Howell, N. Pisias, J. Ballance, J. Baughman and L. Ochs, Brown University), which uses the Blackman–Tukey technique. Phases were calculated by cross spectral analysis with ETP (the sum of normalized eccentricity plus obliquity minus the precession index) using a maximum lag of 150 kyr. Error bars on phase estimates do not include the age model uncertainty of ~4 kyr. This uncertainty affects phases relative to Milankovitch forcing but not relative to benthic $\delta^{18}O$, and may explain why Caribbean SST and minimum $\Delta\delta^{13}C_{mid}$ appear to lead June perihelion slightly.

1. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992).
2. Curry, W. B. in *The South Atlantic: Past and Present Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D.) 577–598 (Springer, 1996).
3. Weber, S. L. *et al.* The modern and glacial overturning circulation in the Atlantic ocean in PMIP coupled model simulations. *Clim. Past* **3**, 51–64 (2007).
4. Gregory, J. M. *et al.* A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration. *Geophys. Res. Lett.* **32**, L12703, doi:10.1029/2005GL023209 (2005).
5. McManus, J. F., Oppo, D. W., Keigwin, L. D., Cullen, J. L. & Bond, G. C. Thermohaline circulation and prolonged interglacial warmth in the North Atlantic. *Quat. Res.* **58**, 17–21 (2002).
6. Boyle, E. A. & Keigwin, L. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
7. McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
8. Dahl, K. A., Broccoli, A. J. & Stouffer, D. J. Assessing the role of North Atlantic freshwater forcing in millennial scale climate variability: A tropical Atlantic perspective. *Clim. Dyn.* **24**, 325–346 (2005).
9. Curry, W. B. & Oppo, D. W. Glacial water mass geometry and the distribution of $\delta^{13}C$ of ΣCO_2 in the western Atlantic Ocean. *Paleoceanography* **20**, PA1017, doi:10.1029/2004PA001021 (2005).
10. Marchitto, T. M. & Broecker, W. S. Deep water mass geometry in the glacial Atlantic Ocean: A review of constraints from the paleonutrient proxy Cd/Ca. *Geochem. Geophys. Geosyst.* **7**, Q12003, doi:10.1029/2006GC001323 (2006).
11. Mackensen, A., Hubberton, H.-W., Bickert, T., Fischer, G. & Fütterer, D. K. The $\delta^{13}C$ in benthic foraminiferal tests of *Fontbotia Wuellerstorfi* (Schwager) relative to the $\delta^{13}C$ of dissolved inorganic carbon in Southern Ocean Deep Water: Implications for glacial ocean circulation models. *Paleoceanography* **8**, 587–610 (1993).
12. McCave, I. N., Manighetti, B. & Beveridge, N. A. S. Circulation in the glacial North Atlantic inferred from grain-size measurements. *Nature* **374**, 149–152 (1995).
13. Raymo, M. E., Ruddiman, W. F., Shackleton, N. J. & Oppo, D. W. Evolution of the Atlantic-Pacific $\delta^{13}C$ gradients over the last 2.5 m.y. *Earth Planet. Sci. Lett.* **93**, 353–368 (1990).
14. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20**, PA1003, doi:10.1029/2004PA001071 (2005).
15. Skinner, L. C. & Shackleton, N. J. An Atlantic lead over Pacific deep-water change across Termination I: Implications for the application of the marine isotope stage stratigraphy. *Quat. Sci. Rev.* **24**, 571–580 (2005).
16. Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Backman, J. Pleistocene evolution: Northern hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography* **4**, 353–412 (1989).
17. Huybers, P., Gebbie, G. & Marchal, O. Can paleoceanographic tracers constrain meridional circulation rates? *J. Phys. Oceanogr.* **37**, 394–407 (2007).
18. Ruhlemann, C., Mulitza, S., Muller, P. J., Wefer, G. & Zahn, R. Warming of the tropical Atlantic Ocean and slowdown of thermohaline circulation during the last deglaciation. *Nature* **402**, 511–514 (1999).
19. Ruddiman, W. F. & McIntyre, A. Oceanic mechanisms for amplification of the 23,000-year ice-volume cycle. *Science* **212**, 617–627 (1981).
20. Imbrie, J., McIntyre, A. & Mix, A. in *Climate and Geo-Sciences* (eds Berger, A. *et al.*) 121–164 (Kluwer Academic, 1989).
21. Schmidt, M. W., Vautravers, M. J. & Spero, H. J. Western Caribbean sea surface temperatures during the late Quaternary. *Geochem. Geophys. Geosyst.* **7**, Q02P10, doi:10.1029/2005GC000957 (2006).
22. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
23. Wang, Z. & Mysak, L. A. Simulation of the last glacial inception and rapid ice sheet growth in the McGill paleoclimate model. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL015120 (2002).
24. Khodri, M. *et al.* Modelling the climate evolution from the last interglacial to the start of the last glaciation: The role of Arctic Ocean freshwater budget. *Geophys. Res. Lett.* **30**, 1606, doi:10.1029/2003GL017108 (2003).
25. Yoshimori, M., Weaver, A. J., Marshall, S. J. & Clarke, G. K. C. Glacial termination: Sensitivity to orbital and CO_2 forcing in a coupled climate system model. *Clim. Dyn.* **17**, 571–588 (2001).
26. Crucifix, M. & Loutre, M. F. Transient simulations over the last interglacial period (125–115 kyr BP): Feedback and forcing analysis. *Clim. Dyn.* **19**, 417–433 (2002).
27. Tuenter, E., Weber, S. L., Hilgen, F. J., Lourens, L. J. & Ganopolski, A. Simulation of climate phase lags in response to precession and obliquity forcing and the role of vegetation. *Clim. Dyn.* **24**, 279–295 (2005).
28. Shin, S., Liu, Z., Otto-Bliesner, B., Brady, E. & Kutzbach, J. Southern Ocean sea-ice control of the glacial North Atlantic thermohaline circulation. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL015513 (2003).
29. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–917 (2007).
30. Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313**, 508–511 (2006).

Author Contributions L.E.L. designed the study, performed the spectral analysis, and wrote the paper in consultation with M.E.R. and W.B.C.; contributed unpublished data from ODP Sites 926 and 928.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.E.L. (lisiecki@alumni.brown.edu).

LETTERS

The Earth's missing lead may not be in the core

M. Lagos[1,2], C. Ballhaus[1], C. Münker[1,2], C. Wohlgemuth–Ueberwasser[1,2], J. Berndt[2] & Dmitry V. Kuzmin[3,4]

Relative to the CI chondrite class of meteorites (widely thought to be the 'building blocks' of the terrestrial planets), the Earth is depleted in volatile elements. For most elements this depletion is thought to be a solar nebular signature, as chondrites show depletions qualitatively similar to that of the Earth[1]. On the other hand, as lead is a volatile element, some Pb may also have been lost after accretion. The unique $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ ratios of the Earth's mantle suggest that some lead was lost about 50 to 130 Myr after Solar System formation[2–4]. This has commonly been explained by lead lost via the segregation of a sulphide melt to the Earth's core[5–7], which assumes that lead has an affinity towards sulphide. Some models, however, have reconciled the Earth's lead deficit with volatilization[8]. Whichever model is preferred, the broad coincidence of U–Pb model ages with the age of the Moon[9–11] suggests that lead loss may be related to the Moon-forming impact. Here we report partitioning experiments in metal–sulphide–silicate systems. We show that lead is neither siderophile nor chalcophile enough to explain the high U/Pb ratio of the Earth's mantle as being a result of lead pumping to the core. The Earth may have accreted from initially volatile-depleted material, some lead may have been lost to degassing following the Moon-forming giant impact, or a hidden reservoir exists in the deep mantle with lead isotope compositions complementary to upper-mantle values; it is unlikely though that the missing lead resides in the core.

The separation of metal from silicate is one of the most important events in the Earth's early chemical differentiation. On the basis of Hf–W and U–Pb model ages, it is assumed that core formation progressed in distinct episodes[6,7,12]: first the segregation of a reduced (Fe,Ni) metal melt, and then the segregation of an oxidized FeS sulphide melt. The basis for this assumption is an apparent disparity between the Hf–W and U–Pb model ages of the Earth's mantle[4,6,7]. Tungsten is moderately siderophile and fractionated to the core most efficiently by an (Fe,Ni) metal melt. Tungsten isotope signatures of the Earth's mantle indicate that (Fe,Ni) metal segregation lasted until at least 30 Myr after formation of the Solar System[8,9]. Lead, in contrast, is believed to be chalcophile but not siderophile[1,5,13], and therefore depleted most efficiently by sulphide. The putative sulphide segregation event is dated by U–Pb model ages to ~50–130 Myr[2,4,6] after Solar System formation. Because the U–Pb model ages broadly coincide with the age of the Moon[4,10,11], it is hypothesized[6,7] that sulphide segregation and related lead loss were triggered by the Moon-forming impact.

For the Earth's missing lead to have been sequestered to the core, lead must be sufficiently siderophile and/or chalcophile. Furthermore, if we are to propose that the U–Pb ages date a discrete sulphide segregation event[6,7], it must be demonstrated that after early (Fe,Ni) metal segregation, the mantle still contained enough sulphur to stabilize sulphide. We address these questions with experimental partitioning data of lead and the similarly volatile elements cadmium, zinc, selenium and tellurium between (Fe,Ni) metal, FeS sulphide and basaltic silicate melt under upper-mantle conditions (Methods Summary).

The metal–silicate experiments were performed in graphite-lined platinum capsules at 1,400 °C and 0.5 to 2 GPa, in an oxygen fugacity (f_{O_2}) range relative to the iron–wüstite (IW) buffer from IW − 1.3 to IW − 1.8 (Fig. 1). This is close to the f_{O_2} value at which the Earth's mantle (~8 wt% FeO) is in redox equilibrium with the Fe/Ni ratio (90:5) of the Earth's core. At 1,400 °C and 2 GPa (f_{O_2}=IW−2), the partition coefficients (metal–silicate) are $D_{Pb} = 0.18 \pm 0.04$, $D_{Cd} = 0.5 \pm 0.1$, $D_{Zn} = 0.08 \pm 0.015$, $D_{Se} = 13 \pm 4$ and $D_{Te} = 27 \pm 8$ (uncertainties, 2 s.d.; Supplementary Information). D_{Pb} appears to be pressure dependent. A D_{Pb} (metal–silicate) determined at 20 GPa and IW − 4 (ref. 14) and extrapolated to IW − 2 suggests that with increasing pressure, lead becomes more lithophile[15,16].
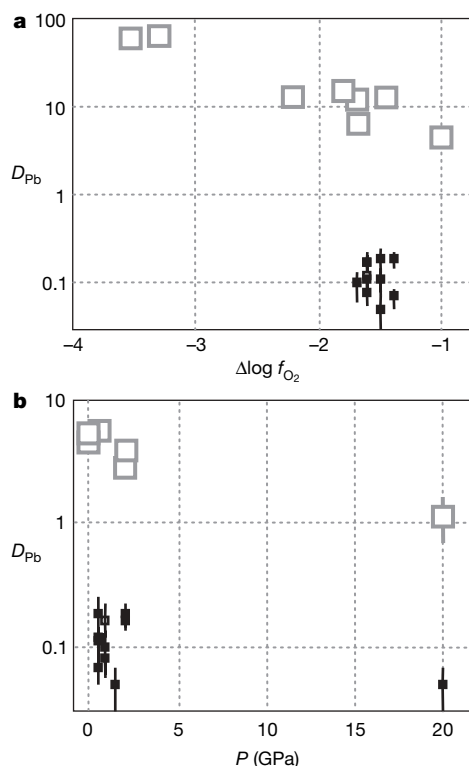

Figure 1 | Metal–silicate (black) and sulphide–silicate (grey) partition coefficients for lead. a, Data as a function of f_{O_2} relative to IW; **b**, data as a function of pressure. In **b**, the D_{Pb} (metal–silicate) values are extrapolated to IW − 2 and the D_{Pb} (sulphide–silicate) values are extrapolated to IW − 1, assuming lead to be divalent in silicate melt (see text) and dissolved by the generalized metal–silicate redox equilibrium
M (metal) + $n/4O_2$ = $MO_{n/2}$ (silicate), where n is the charge of the metal cation in the silicate melt. One metal–silicate experiment at 1 atm ($D_{Pb} = 0.5$) is from ref. 29, and the experiments at 20 GPa are from ref. 15. Errors in relative f_{O_2} values (not shown for clarity) are around ±0.3 log units. Errors in partition coefficients are 2 s.d.

[1]Steinmann-Institut, Universität Bonn, Poppelsdorfer Schloss, 53115 Bonn, Germany. [2]Institut für Mineralogie, Universität Münster, Corrensstrasse 24, 48149 Münster, Germany. [3]Max-Planck-Institut für Chemie, Abteilung Geochemie, 55128 Mainz, Germany. [4]Institute of Geology and Mineralogy SB RAS, Novosibirsk 630090, Russia.

The sulphide–silicate partition coefficients were determined in iron metal capsules at 1,350 to 1,400 °C and at 1 atm to 2 GPa from IW − 3.5 to IW − 1. With increasing oxidation, the metals become distinctly more lithophile (Fig. 1), contrary to assertions in ref. 6, where it is argued that with progressive mantle oxidation lead should become more chalcophile. We report all partition coefficients for a relative f_{O_2} value of IW − 1, acknowledging that sulphide segregation necessitated f_{O_2} conditions more oxidized than in iron metal saturation[6,12]. At 1,400 °C, 2 GPa and IW − 1, the partition coefficients (sulphide–silicate) are $D_{Pb} = 2.6 \pm 0.7$, $D_{Cd} = 2.4 \pm 0.5$, $D_{Zn} = 0.12 \pm 0.02$, $D_{Se} = 5.1 \pm 1.7$ and $D_{Te} = 8.4 \pm 2.3$ (uncertainties, 2 s.d.). D_{Pb} (sulphide–silicate) is pressure sensitive. A sulphide–silicate experiment[14] at 20 GPa, extrapolated to IW − 1 assuming that lead is divalent in silicate melt, shows that with increasing pressure lead becomes more lithophile.

The partition coefficients are summarized in Fig. 2. Lead and, to some extent, cadmium are moderately chalcophile or lithophile but never siderophile; zinc is lithophile in both metal–silicate and sulphide–silicate systems; and selenium and tellurium are both chalcophile and siderophile.

The partition coefficients make it possible to test how well actual mantle abundances of lead, zinc, cadmium, selenium and tellurium can be reproduced by the core-formation scenarios outlined above. Initial element abundances in the bulk Earth are taken from ref. 17. For the first stage, the (Fe,Ni) metal segregation, we assume that during accretion and core segregation the metal/silicate mass ratio (32:68) remained constant, that accretion decreased exponentially with time, and that equilibrium between metal and silicate was maintained throughout core-melt segregation[8].

Overall, the match with mantle abundances[1] is quite satisfactory (Fig. 3a). Zinc, cadmium and, particularly, lead appear to be elevated in the model by a factor of ~2; however, if early (Fe,Ni) metal were to segregate at conditions more reduced than IW − 2 (ref. 12), then the match would quickly improve, because as f_{O_2} decreases metals become more siderophile. Modelled abundances of selenium and tellurium are apparently slightly below mantle abundances, but we note that selenium and tellurium abundances in the Earth's mantle[1] were inferred from sulphur abundances.

To model the second-stage, sulphide segregation, event (Fig. 3b), we assume that all sulphide that was segregated to the core came from the impactor. Because sulphur is siderophile[7], the proto-Earth's mantle after segregation of early (Fe,Ni) metal melt must have been essentially sulphur free. Moreover, we assume that the impactor was as volatile depleted as the Earth, that is, that it had a sulphur content of 6,350 p.p.m. (ref. 1). As an extra 10% of material was added to the proto-Earth during the Moon-forming impact[6,7,12], only about 0.25%

of the FeS could have been segregated. Given such a small amount, batch segregation is assumed.

After the giant impact and putative sulphide segregation, agreement between the model and the natural mantle deteriorates (Fig. 3b). The modelled selenium and tellurium abundances increase relative to first-stage (Fe,Ni) metal segregation by a factor of ten. If the impactor had to replenish the proto-Earth's mantle with sulphur to allow for subsequent sulphide segregation, it would have added corresponding amounts of selenium and tellurium, which were difficult to remove subsequently by segregating sulphide. For the Pb–Cd–Zn abundances of the Earth's mantle, the impactor has no effect resolvable on the scale of Fig. 3b, because the partition coefficients (sulphide–silicate) are too low and the amount of sulphide segregated is too small. Hence, abundances in chalcophile elements do not seem to require a discrete sulphide segregation event. This point is emphasized further with lead isotopes.

Figure 4a depicts the result of a reverse model in which we calculate the $^{238}U/^{204}Pb$ ratio of the bulk Earth at the time of Solar System formation, using a reasonable present-day $^{238}U/^{204}Pb$ ratio (μ) for the bulk silicate Earth of $\mu = 8$ (refs 2, 4, 18). Fifty million years after Solar System formation, that is, the age assumed for the giant impact[10,11], the silicate Earth's $^{238}U/^{204}Pb$ ratio was ~16. The putative sulphide segregation event has no significant effect on $^{238}U/^{204}Pb$ because D_{Pb} (sulphide–silicate) is too low and/or the amount of sulphide (0.25%) segregated is too small. Recalculating the effect of (Fe,Ni) metal segregation back to the beginning of the Earth's accretion, we obtain a $^{238}U/^{204}Pb$ ratio for the bulk Earth at time zero of between 15 (metal segregation at IW − 2) and 10 (metal segregation at IW − 4). Reasonable estimates of the bulk Earth's $^{238}U/^{204}Pb$ ratio at that time, however, are around 1.4 (ref. 18), which is almost an order of magnitude lower than the $^{238}U/^{204}Pb$ ratio range calculated here. For comparison, the $^{238}U/^{204}Pb$ ratios of CI and CV chondrites 4.56 Gyr ago were ~0.2 and 1 (refs 18, 19), respectively.

In Fig. 4b we attempt to reproduce in a forward model the μ value of the present-day upper mantle, using a value of 1.4 for the bulk
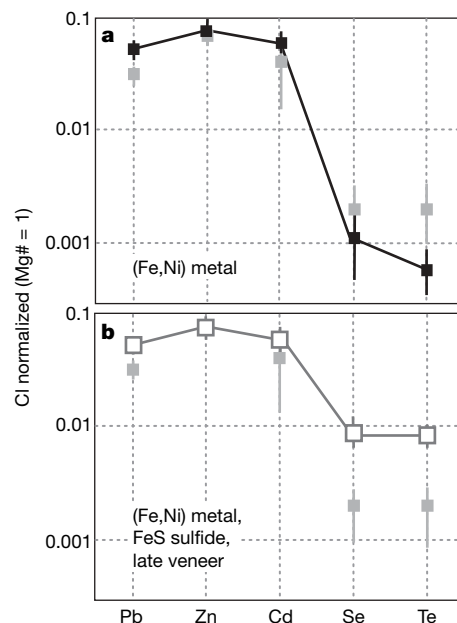
Figure 3 | Calculation of lead, cadmium, zinc, selenium and tellurium abundances in the Earth's mantle with partition coefficients summarized in Fig. 2, normalized to CI and Mg# = 1. a, The mantle after 89% accretion[12] and simultaneous segregation of an (Fe,Ni) metal melt at an f_{O_2} value of IW − 2. **b**, The mantle after 10% material addition by the Moon-forming impactor, 0.25% sulphide segregation, and finally the addition of 1% of a late chondritic veneer (see text). Superimposed in lighter grey are mantle abundances from ref. 1. Error bars, 2 s.d.
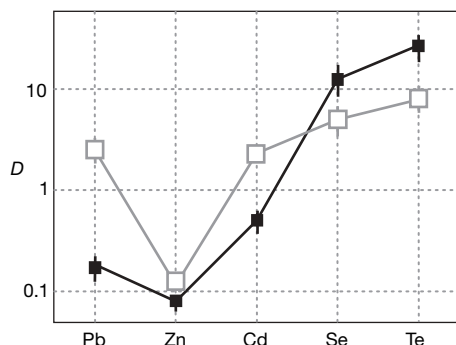

Figure 2 | Summary of the metal–silicate (black) and sulphide–silicate (grey) partition coefficients (1,400 °C, 2 GPa). The metal–silicate partition coefficients are reported for a relative f_{O_2} value of IW − 2, and the sulphide–silicate partition coefficients for IW − 1, assuming that segregation of a FeS melt requires f_{O_2} conditions more oxidized than in iron metal saturation. Error bars, 2 s.d.

^{238}U/^{204}Pb ratio 4.56 Gyr ago. This model yields a μ value for the Earth's mantle of ~1. Again, this is an order of magnitude lower than μ values deemed reasonable for the Earth's mantle. Figure 4c illustrates how large D_{Pb} (sulphide–silicate) must be—on the order of 1,000—for lead to be sequestered to the core. No experimental evidence exists to support a D_{Pb} value of that magnitude.

It seems that the Earth's high U/Pb ratio cannot be explained easily by lead pumping to the core. Alternatives are that the Earth accreted from initially volatile-depleted material[20], or that some lead was lost to degassing following the giant impact. Accretion from volatile-depleted material alone cannot explain all the features observed, despite that fact that, in terms of volatile-depletion pattern, the Earth is remarkably similar to carbonaceous chondrites[1]. The silicate Earth's unique lead isotope composition (Fig. 5) requires that at least some lead loss occurred after accretion. Degassing to a hot 'silicate atmosphere' from a magma ocean[21] coupled with atmospheric loss through impact erosion may be an alternative explanation, because U–Pb model ages overlap with closure ages of the Earth's atmosphere obtained with the I–Pu–Xe chronometers[4,22–24] (~100 Myr after Solar System formation). A third alternative may be that the lead isotope
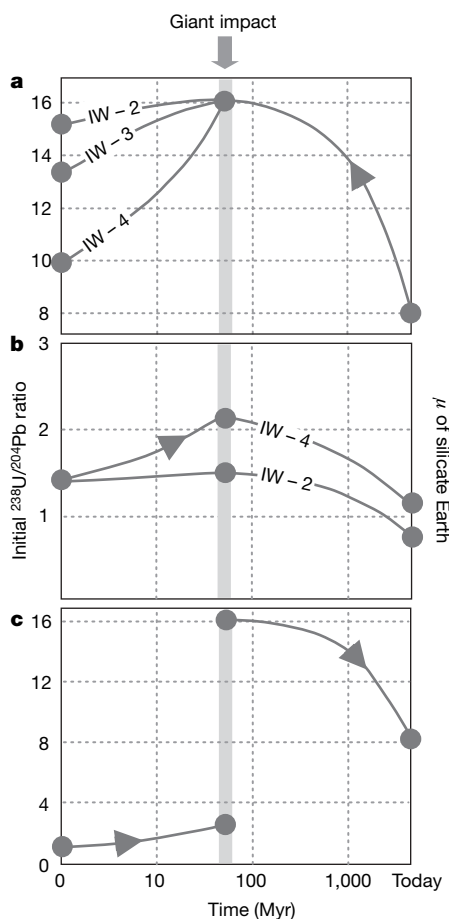

Figure 5 | Modelled lead isotope compositions of the present-day upper mantle. Curves 1 and 2 illustrate the effect on lead isotope ratios of (Fe,Ni) core-melt segregation at relative f_{O_2} conditions of IW − 2 (curve 1) and IW − 4 (curve 2); regardless of f_{O_2}, no significant displacement to the right of the geochron is achieved. Curve 3 describes the lead isotope evolution of the upper mantle with an initial ^{238}U/^{204}Pb ratio of 1.4 (ref. 18) and (Fe,Ni) melt segregation at IW − 2, followed by lead evaporation at 50 Myr (giant impact) to an extent that the present-day condition $\mu = 8$ is satisfied. This model accounts for the ^{206}Pb/^{204}Pb and ^{207}Pb/^{204}Pb isotope ratios of average mid-ocean-ridge basalt[3].

composition of the bulk silicate Earth is different from the presently estimated value inferred from oceanic basalts. It is possible that a hidden reservoir exists in the deep mantle with lead isotope compositions complementary to upper-mantle values[24].

In summary, the bulk abundances of lead, cadmium, zinc, selenium and tellurium, the ^{238}U/^{204}Pb ratio and the Pb isotope ratios of the Earth's mantle cannot be reconciled with a sulphide segregation event during core formation. Any lead in the Earth's core must have been sequestered there by means of an early (Fe,Ni) metal melt. However, owing to the lithophile character of lead, the effect on the ^{206}Pb/^{204}Pb and ^{207}Pb/^{204}Pb ratios of the Earth's mantle is likely to be minimal (Fig. 5). We may even question whether the proto-Earth's mantle could have been replenished with sulphur during the giant impact. A 50-Myr-old, Mars-sized protoplanet was almost certainly differentiated into silicate and metal[9], so its sulphur content would have been contained in its core, unlikely to be released from metal to the silicate mantle after the impact. Furthermore, a FeS melt is unlikely to have been stable. A fertile mantle composition, when pressurized to depths of over 250 km, becomes self-reduced to ~IW − 2, at which point a FeS melt is converted to an iron metal melt with sulphur as its minor component[25]. If IW − 2 is the prevailing redox state of the present deeper mantle[25–27], it is unreasonable to expect more oxidized conditions at Hadean times.

METHODS SUMMARY

Experiments were performed at 1,350 to 1,400 °C from 1 atm to 2 GPa (Supplementary Table 1). The metal–silicate experiments used welded platinum tubes with an inner graphite lining. The sulphide–silicate experiments were performed in iron metal capsules. Starting materials were a natural basalt with ~5.8 wt% MgO and 6 wt% FeO (sample S40 in ref. 28), iron metal and synthetic FeS. Lead, zinc, cadmium, selenium and tellurium were added as oxides in percentage to hundreds-of-parts-per-million concentrations (Supplementary Table 1). Oxygen fugacity (f_{O_2}) is given by $\Delta \log f_{O_2} = 2\log(X_{FeO}^{silicate}/X_{Fe}^{metal})$ relative to the IW buffer, where X is the mole fraction of the component in the phase. Conditions more reduced than IW − 2 were achieved by adding traces of metallic silicon. Sulphur fugacity (f_{S_2}) is a function of the metal/sulphur ratio of the sulphide phase. Products are silicate glass, iron metal and Fe_{1+x}S sulphide.

Major elements were determined using electron-probe microanalysis (Supplementary Table 1). Lead, cadmium, zinc, selenium and tellurium were analysed with a 193-nm ArF excimer laser coupled to a magnetic sector ICP mass spectrometer. Laser beam diameters were around 60 μm and laser energies ranged from 4 J cm^{-2} (metal and sulphide) to 8 J cm^{-2} (silicate). Isotopes quantified were ^{57}Fe, ^{64}Zn, ^{66}Zn, ^{67}Zn, ^{68}Zn, ^{110}Cd, ^{111}Cd, ^{112}Cd, ^{113}Cd, ^{114}Cd, ^{116}Cd, ^{77}Se, ^{82}Se, ^{125}Te, ^{126}Te, ^{128}Te, ^{130}Te, ^{206}Pb, ^{207}Pb and ^{208}Pb. No isobaric interferences were noted. Partition coefficients calculated with different isotopes of the same

Figure 4 | Modelled evolution of the silicate Earth's ^{238}U/^{204}Pb ratio in time elapsed after Solar System formation, using the partition coefficients summarized in Fig. 2. Core-formation schemes as described in the text, assuming (Fe,Ni) metal melt segregation until 50 Myr after Solar System formation, followed at that time by a sulphide segregation event. Calculations were made for IW − 2 to IW − 4 to accommodate uncertainties in relative f_{O_2} values during early (Fe,Ni) metal segregation[12]. **a**, Reverse model, calculating the initial bulk Earth ^{238}U/^{204}Pb ratio assuming a present day ^{238}U/^{204}Pb ratio for bulk silicate Earth of $\mu = 8$. **b**, Forward model, calculating the μ value of the bulk silicate Earth starting with a reasonable bulk Earth ^{238}U/^{204}Pb ratio estimate of 1.4 (ref. 18) at the time of Solar System formation. **c**, Forward and reverse model combined; a match can be achieved with a D_{Pb} (sulphide–silicate) value >1,000, for which there is no experimental evidence.
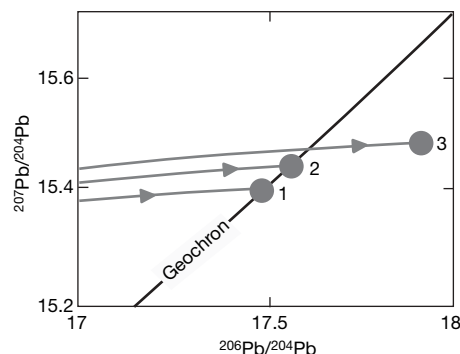
91

element were identical within their uncertainties. All count rates were background corrected, then normalized to the ^{57}Fe count rates and the total iron content of the phase. The metal–silicate and sulphide–silicate partition coefficients (Supplementary Table 2) were derived by dividing normalized count rates of coexisting metal–silicate and sulphide–silicate pairs.

1. Palme, H. & O'Neill, H. S. in *The Mantle and Core* (ed. Carlson, R. W.) 1–38 (Treatise on Geochemistry 2, Elsevier, 2003).
2. Galer, S. J. G. & Goldstein, S. L. in *Earth Processes: Reading the Isotope Code* (eds Basu, A. & Hart, S.) 75–98 (American Geophysical Union, 1996).
3. Hofmann, A. W. in *The Mantle and Core* (ed. Carlson, R. W.) 61–101 (Treatise on Geochemistry 2, Elsevier, 2003).
4. Allègre, C. J., Manhès, G. & Göpel, C. The major differentiation of the Earth at ∼ 4.45 Ga. *Earth Planet. Sci. Lett.* **267,** 386–398 (2008).
5. Oversby, V. M. & Ringwood, A. E. Time of formation of the Earth's core. *Nature* **234,** 463–465 (1971).
6. Wood, B. J. & Halliday, A. N. Cooling of the Earth and core formation after the giant impact. *Nature* **437,** 1345–1348 (2005).
7. Wood, B. J., Walter, M. J. & Wade, J. Accretion of the Earth and segregation of its core. *Nature* **441,** 825–833 (2006).
8. Jacobsen, S. B. The Hf-W isotopic system and the origin of the earth and moon. *Annu. Rev. Earth Planet. Sci.* **33,** 531–570 (2005).
9. Kleine, T., Münker, C., Mezger, K. & Palme, H. Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* **418,** 952–955 (2002).
10. Kleine, T. *et al.* Hf-W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* **310,** 1671–1674 (2005).
11. Touboul, M. *et al.* Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* **450,** 1206–1209 (2007).
12. O'Neill, H. & St C.. The origin of the Moon and the early history of the Earth - A chemical model. Part 2: The Earth. *Geochim. Cosmochim. Acta* **55,** 1159–1172 (1991).
13. Hart, S. R. & Gaetani, R. A. Mantle Pb paradoxes: the sulfide solution. *Contrib. Mineral. Petrol.* **152,** 295–308 (2006).
14. Malavergne, V. *et al.* New high-pressure and high-temperature metal/silicate partitioning of U and Pb: Implications for the cores of the Earth and Mars. *Geochim. Cosmochim. Acta* **71,** 2637–2655 (2007).
15. Petitet, J. P., Petot, C. & Petot-Ervas, G. Influence of pressure on the activity of PbO in an equimolar molten PbO-SiO$_2$ mixture. *Chem. Geol.* **62,** 31–34 (1987).
16. Ohtani, E., Yurimoto, H. & Seto, S. Element partitioning between metallic liquid, silicate liquid, and lower-mantle minerals: implications for core formation of the Earth. *Phys. Earth Planet. Inter.* **100,** 97–114 (1997).
17. McDonough, W. F. in *The Mantle and Core* (ed. Carlson, R. W.) 547–568 (Treatise on Geochemistry 2, Elsevier, 2003).
18. Allegre, C. J., Manhes, G. & Göpel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59,** 1445–1456 (1995).
19. Palme, H. & Jones, A. in *Meteorites, Comets and Planets* (ed. Davis A. M.) 41–61 (Treatise on Geochemistry 1, Elsevier, 2003).
20. Yin, Q.-Z. & Jacobsen, S. B. Does U–Pb date Earth's core formation? *Nature* **444,** E1, doi:10.1038/nature05358 (2006).
21. Solomatov, V. S. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 323–338 (Univ. Arizona Press, 2000).
22. Porcelli, D., Woolum, D. & Cassen, P. Deep Earth rare gases: initial inventories, capture from the solar nebula, and losses during Moon formation. *Earth Planet. Sci. Lett.* **193,** 237–251 (2001).
23. Ozima, H. & Podosek, F. A. Formation age of the earth from ^{129}I/^{127}I and ^{244}Pu/^{238}U systematics and the missing Xe. *J. Geophys. Res.* **104,** 25493–25499 (1999).
24. Tolstikhin, I. & Hofmann, A. W. Early crust on top of the Earth's core. *Phys. Earth Planet. Inter.* **148,** 109–130 (2005).
25. Rohrbach, A. *et al.* Metal saturation in the upper mantle. *Nature* **449,** 456–458 (2007).
26. O'Neill, H. & St. C. . *et al.* Mössbauer spectroscopy of mantle transition zone phases and determination of minimum Fe3+ content. *Am. Mineral.* **78,** 456–460 (1993).
27. Frost, D. J. *et al.* Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle. *Nature* **428,** 409–412 (2004).
28. Schuth, S. *et al.* Geochemical constraints on the petrogenesis of arc picrites and basalts, New Georgia Group, Solomon Islands. *Contrib. Mineral. Petrol.* **148,** 288–304 (2004).
29. Jones, J. H. & Drake, M. J. Geochemical constraints on core formation in the Earth. *Nature* **322,** 221–228 (1986).

LETTERS

Linking climate change to lemming cycles

Kyrre L. Kausrud[1], Atle Mysterud[1], Harald Steen[2]†, Jon Olav Vik[1], Eivind Østbye[2], Bernard Cazelles[3,4], Erik Framstad[5], Anne Maria Eikeset[1], Ivar Mysterud[2], Torstein Solhøy[6] & Nils Chr. Stenseth[1]

The population cycles of rodents at northern latitudes have puzzled people for centuries[1,2], and their impact is manifest throughout the alpine ecosystem[2,3]. Climate change is known to be able to drive animal population dynamics between stable and cyclic phases[4,5], and has been suggested to cause the recent changes in cyclic dynamics of rodents and their predators[3,6–9]. But although predator–rodent interactions are commonly argued to be the cause of the Fennoscandian rodent cycles[1,10–13], the role of the environment in the modulation of such dynamics is often poorly understood in natural systems[8,9,14]. Hence, quantitative links between climate-driven processes and rodent dynamics have so far been lacking. Here we show that winter weather and snow conditions, together with density dependence in the net population growth rate, account for the observed population dynamics of the rodent community dominated by lemmings (*Lemmus lemmus*) in an alpine Norwegian core habitat between 1970 and 1997, and predict the observed absence of rodent peak years after 1994. These local rodent dynamics are coherent with alpine bird dynamics both locally and over all of southern Norway, consistent with the influence of large-scale fluctuations in winter conditions. The relationship between commonly available meteorological data and snow conditions indicates that changes in temperature and humidity, and thus conditions in the subnivean space, seem to markedly affect the dynamics of alpine rodents and their linked groups. The pattern of less regular rodent peaks, and corresponding changes in the overall dynamics of the alpine ecosystem, thus seems likely to prevail over a growing area under projected climate change.

Winter conditions are likely to be critical for the demography of many high-latitude rodents[7,15,16]. When available, the subnivean space provides thermal insulation, access to food plants and protection from generalist predators like foxes, owls, corvids and raptors[2,16–18]. Norway lemmings and several other Fennoscandian rodents will even commence reproduction in the subnivean if conditions are favourable[2,18]. Changes in the condition and/or duration of the subnivean habitat are thus likely to affect the performance of the rodent community through temperature stress, flooding risk, food limitation and even predator access[2,8,16–20].

Here we combine long-term field estimates of snow conditions with meteorological data to estimate the effect of winter weather fluctuations on snow conditions. Using a 38-year record of rodent trap data (Fig. 1a), we then estimate the effects of snow conditions (Fig. 2a) on the dynamics of the alpine rodent community (Fig. 3), focusing on the numerically dominant lemmings. Using censuses of the local ground-nesting bird communities as well as large-scale data from the annual ptarmigan and willow grouse hunting season, we also assess whether such effects are being transmitted to rodent-linked communities on local and/or regional scales. Wavelet analyses (Fig. 1c–e, Supplementary Figs 10, 11, 14) confirm that all rodents

and birds within our study area had a 3–5-year dominant period in the 1970s and 1980s (that is, before a period of recent warming; see Supplementary Figs 6, 7). The dynamics of both lemmings and other rodents, as well as of the ptarmigan/willow grouse, changed as cyclicity faded in the late 1990s (Fig. 1). With fading cycles, the coherence between lemmings and other rodents abundances also disappeared (1970–1995: $r = 0.70$, $n = 49$, $P < 0.01$; 1996–2007: $r < 0.02$, $n = 22$, $P > 0.50$; see also Supplementary Fig. 14).

Because the formation of subnivean space produces snow crystals with weak cohesion near the ground, the hardness of the bottom of the snowpack is often a good indicator of subnivean conditions[19]. In 15 of the years 1970–2007, this was measured using snow wells dug in late winter (see Methods). The mean measurement is closely negatively correlated with the logarithmic rate of change in total rodent abundance from one spring to the next ($r = -0.80$, $n = 15$, $P > 0.01$; Fig. 2b). The mean number of crusts in the snowpack is closely correlated with the mean measured ground snow hardness ($r = 0.71$, $n = 11$, $P < 0.01$), pointing to the latter being an effect of temperature fluctuations. Indeed, we found that snow hardness for the other 23 years could be predicted from the temperature fluctuations throughout winter (see equations (3) and (4), in Methods), explaining 68% of the observed variance. This predicted hardness was then found to be almost as closely correlated with rodent abundance change over winter ($r = -0.66$, $n = 22$, $P < 0.01$). This is supported by recent experimental evidence that extension of the available subnivean space increases winter survival of the root vole (*Microtus oeconomus*)[16].

Relative air humidity probably reflects significant differences in the amount of free water, and is thus related to heat loss and risk of flooding as well as ice formation[18–20]. This is likely to be important for newborn and lactating females in the subnivean space[8,18]. Indeed, in some winters there appear to have been sizeable populations in late winter that collapsed before spring trapping (E. F., unpublished observations), suggesting a critical spring phase. Rodent abundances—but not rates of change over winter—correlate negatively with relative humidity in April ($r = -0.52$, $n = 24$, $P < 0.01$) measured at Finse meteorological station. It has been suggested[2] that successful spring reproductive phases for the rodent species that start reproducing under the snow contribute to high summer peaks by swamping generalist predators. We modelled fourteen years (1991, 1995–2007) of humidity data using temperature and precipitation (see equation (5), in Methods), and found that they explained 74% of the observed variance. Most of the negative correlation between April humidity and rodent density stems from the fact that five of the six rodent peak years for which humidity measurements are available had median relative humidities of less than 81%, and values this low are predicted not to have occurred since 1996 (Fig. 2). Indeed, a significantly higher median April humidity is predicted after this time (difference between the

---

[1]Centre for Ecological and Evolutionary Synthesis, [2]Department of Biology, University of Oslo, PO Box 1066 Blindern, N-0316 Oslo, Norway. [3]CNRS UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris, France. [4]IRD GEODES, 32 Avenue Henri Varagnat, 93142 Bondy cedex, France. [5]Norwegian Institute for Nature Research, Gaustadalleen 21, N-0349 Oslo, Norway. [6]Department of Biology, University of Bergen, Realfagbygget, Allegaten 41, N-5007 Bergen, Norway. †Present address: Norwegian Polar Institute, N- 9296 Tromsø, Norway.
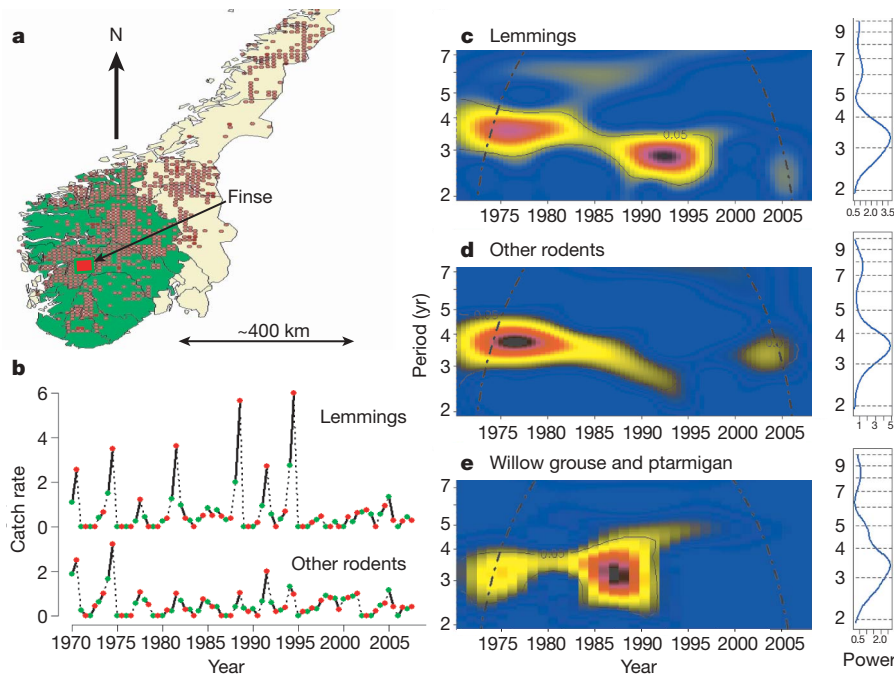
Figure 1 | Population time series. a, Map of South Norway, showing lemming distribution (brown; see http://www.zoologi.no/patlas/kart/ lemen.gif), Finse (red) and counties overlapping the central massif (green). **b**, The rodent catch rates at Finse (green, spring; red, fall; for clarity, we display the square roots of the data). All catch rates are expressed as number caught per 100 trap nights. **c–e**, Wavelet power spectra showing the periodicity of the Finse lemmings (**c**) and other rodents (**d**), and logarithmic rate of change in the ptarmigan and willow grouse hunting returns over the counties highlighted in **a** (**e**). Shifts in periodicity are evident inside the 95% confidence areas (solid black line) and cone of influence (broken black line) (see Methods). Time-averaged spectra show the dominance of the three- to four-year period.

1970–1997 mean and the 1998–2007 mean, 5 percentage points; 18 degrees of freedom, $P > 0.001$). Accordingly, the correlation between humidity and rodent abundance disappears after the last rodent peak of 1994.

To look for effects of the duration and magnitude of snow cover *per se*, we used field estimates of the percentage of ground still covered by snow in mid-July. We found that the North Atlantic Oscillation (NAO; see Methods and Supplementary Information),
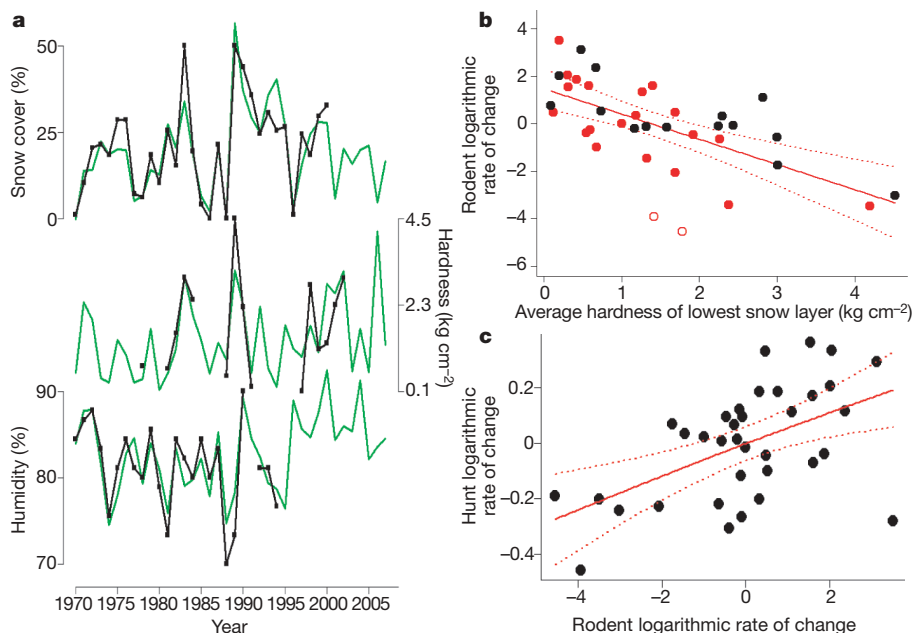

Figure 2 | Climate. a, Data (black) and modelled proxies (green) for the environmental variables found to affect rodent dynamics. **b**, Logarithmic rate of change in rodent abundance, plotted against ground snow hardness. The relationship holds both for observations (black) and the independently modelled proxies based on winter climate (red). The data for 1974–1975 and 1994–1995 (open circles) have the highest rodent populations in the first spring, so these two points slightly lower (but parallel) to the others are expected from predator responses. **c**, Logarithmic rate of change in rodent catch rates at Finse, plotted against the logarithmic rate of change in South Norway ptarmigan and willow grouse hunting returns (Fig. 1 and Supplementary Information).

together with the mean temperatures for October, May and June, explains about 87% of the observed variance (see equation (6), in Methods). Also, the predicted duration of snow cover was not found to be correlated with rodent population growth, but was still found to be a moderately significant explanatory variable in rodent population models (equations 1 and 2).

As global temperatures are expected to rise, we note that temperature is a highly significant predictor of hardness, humidity and duration of the snow cover.

Although spring and autumn densities of rodents are closely correlated ($r = 0.74$, $n = 38$, $P < 0.01$), spring density is linearly independent of the preceding autumn ($r = 0.09$, $n = 38$, $P > 0.50$). Thus, events between autumn and spring seem to be key to predicting between-year fluctuations. By incorporating winter conditions into statistical population models (see equations (1) and (2) and Fig. 3; Methods and Supplementary Information) for the rodent abundance dynamics between 1970 and 1997, we observe that humidity and hardness seem to have strong effects on the over-winter abundance trajectory: together with the previous year's rodent abundances, they are capable of explaining the spring catch rates. The duration of snow cover has considerably less effect (Fig. 3). The autumn abundances, on the other hand, are usually well explained by the spring abundances, with less direct impact from winter conditions.

Despite having predominantly stable mean-field equilibrium, the dynamic behaviour of the models (equations (1) and (2) under environmental stochasticity are consistent with 'cycles' of three to five or more years). This may reconcile the traditional view of rodent fluctuations as limit cycles with the seemingly chaotic dynamics exhibited by several lemming populations[21,22]. The stochastic dynamics captured by our models (see Fig. 3 and Methods) show that the frequency distributions of winter weather variables profoundly influence dynamics without invoking values beyond the observed range. Skewing the distributions of hardness and/or humidity towards increasing values changed the dynamics from three- to five-year cycles towards less frequent peaks and predominantly low-amplitude fluctuations (Fig. 3). The effect of snow duration on cyclicity seemed markedly lower, consistent with the Fennoscandian rodents exhibiting cyclical

tendencies and responses to changing snow conditions over a wide range of altitudes and, thus, snow cover durations.

Notably, the predicted dynamical behaviour emerges from models trained only on 1970–1997 population data. Thus, our predictions do not derive simply from contrasting climate before and after the dynamical shift in the late 1990s, but predicts the absence of rodent peaks after 1994 from the behaviour of the system up to that point.

The logarithmic rate of density change in the local passerine and wader communities (see Methods) are highly correlated with the logarithmic rate of change in rodent density from one spring to the next ($r = 0.69$, $n = 15$, $P < 0.01$ for rodents versus passerines; $r = 0.64$, $n = 14$, $P = 0.01$ for rodents versus waders). Although the ptarmigan and willow grouse data (see Fig. 1 and Methods) was gathered on a much larger spatial scale than the rodent data, there is a high correlation between the logarithmic rate of change in annual rodent abundance at Finse and the logarithmic rate of change in hunting success in the counties overlapping the Hardangervidda massif ($r = 0.65$, $n = 35$, $P < 0.01$; Fig. 1). This correlation stays constant over time, and is reflected in the transition from a three-year period to aperiodicity in the ptarmigan and willow grouse time series in the early 1990s (Fig. 1e). Detrending the ptarmigan and willow grouse data (see Methods and Supplementary Information), we moreover find support for the old observation that there is a positive correlation between the ptarmigan/willow grouse and rodent densities ($r = 0.64$, $n = 36$, $P < 0.01$), even on these different scales. Analysing the counties separately reveals the same pattern (Supplementary Table 5 and Supplementary Fig. 14).

The strong correlations between the annual growth rates of the rodent and different bird communities are consistent with shared predators being an important part of the cyclic and synchronous behaviour of the system[11,18,23], although snow hardness may also have a direct effect on ptarmigan and willow grouse (see Supplementary Fig. 9). Modelling lemmings and other species separately supports the idea that the negative density-dependence term should include all rodent species, despite their different food niches, probably because the reproductive success of many predators depends closely on total rodent abundance[2,6] (even though other agents, like diseases[24,25], may
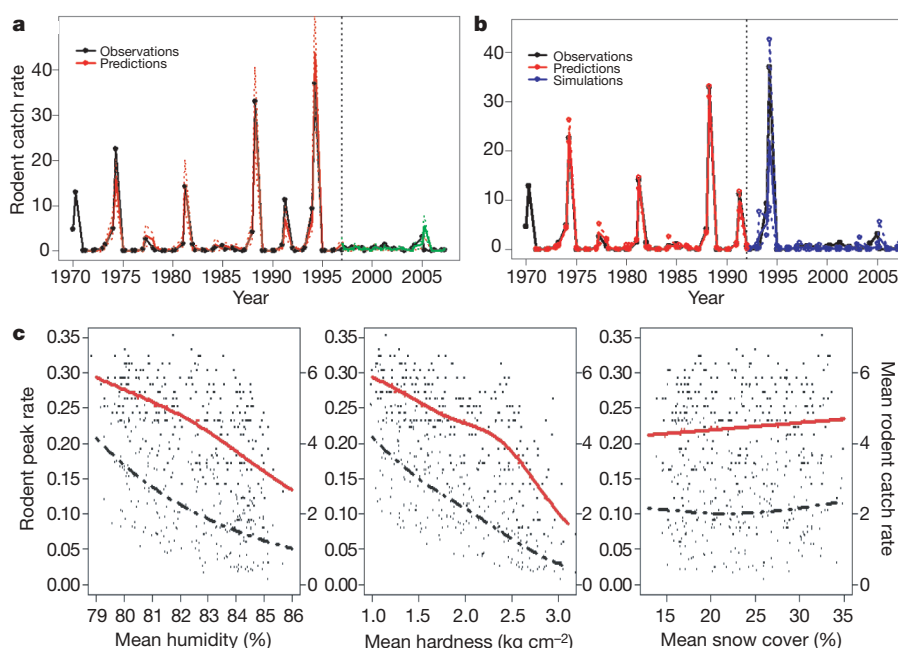

Figure 3 | Models. a, Rodent catch rates with population models (equations (1) and (2)) trained on 1970–1997 data. Black, observations; red, fitted values; green, predictions for 1998–2007. Broken lines indicate the 95% confidence interval of the fit. **b**, As in **a**, but with models trained on 1970–1992 data and the population trajectory simulated for 1992–2008, using (proxy) climate data (blue). The 1994 peak and subsequent absence of peaks are captured. In **b**, the dotted blue line indicates the 95th percentile from 10^3 stochastic simulations. **c**, The climate effects captured by the model: mean rodent peak-year frequencies (red) and mean catch rates (black, small dots) from 10^3 simulations skewing one of the frequency distribution of an environmental variable.

also be involved). The effect seems consistent with the numerical response curve of stoats (*Mustela erminea*) estimated in ref. 26. Specialist predators like stoat and the least weasel (*Mustela nivalis*) can be efficient predators under the snow[16,27] and have highly adapted reproduction strategies tying the number of offspring closely to prey abundance, giving a strongly nonlinear numerical response[26,27]. Although the issue is still debated[28], their numerical response is probably a key causal link between rodent demography and system dynamics.

The abundance relationship between lemmings and the other rodent species suggests that lemming numerical dominance is a result of the extreme peaks, when the lemmings seem to out-reproduce all other species under ideal winter conditions (the correlation between lemming proportion and total catch rate is $r = 0.41$, with $n = 38$, $P < 0.05$). This is responsible for the negative correlation between snow hardness and the proportion of lemmings in the total catch rate ($r = -0.41$, $n = 38$, $P < 0.05$). Lemmings are well known to have very low low-phase population densities, so it is reasonable to expect[4] a decreasing proportion of lemmings in the rodent community when winter conditions remain adverse over time.

The large-scale coherence (see Supplementary Fig. 14) between ptarmigan/willow grouse and rodents is consistent with the considerable spatial autocorrelation in the climate effects, which should have a partial, probabilistic, phase-locking effect on rodent populations over a large area, with corresponding effects on predator-linked species like ptarmigan and willow grouse. However, we expect this to decouple as deteriorating winter conditions decreases the probability of rodent (sub)populations peaking, resulting in less frequent, more local rodent years and correspondingly less potent 'predator pulses' to structure the alpine food web dynamics in space and time. These findings seem consistent with observed spatial and temporal gradients in rodent dynamics, and with the hypothesis[8,21] that snow cover influences the interaction between rodents and specialist (mustelid) versus generalist predators, but indicate that the dynamical effects of predation are dependent on climate-linked processes (see Fig. 4).

Climate reconstructions suggest that the increasingly warm late winter/early spring periods in southeastern Norway over the last decades are unprecedented since 1756[29], when records began. Ongoing climate change may bring more precipitation and higher temperatures[30], and thus probably increase humidity and hard snow over the Scandinavian peninsula, which again will cause the lemming cycle to cease. We can currently only speculate that the absence of occasional or periodic extreme rodent grazing will affect the competitive balance of functional plant groups, with subsequent changes in nutrient cycling. But considering the likely importance of resource pulses for persistence in a poor environment[4], it is probable that the absence of regularly occurring large-scale rodent peak years is responsible for the dramatic declines in arctic foxes and snowy owls on the Scandinavia peninsula[3,9]. On a general level, this points to the fact that environmental changes may perturb any system away from the range of conditions over which it is cyclic. Also, in so far as many naturally occurring cycles involve specialist interactions, which may take time to adjust by migration, demography and/or evolution as communities change, new cycles may appear at a slower rate when the environment changes as quickly as currently seems to be the case.

METHODS SUMMARY

The observed catch rate, $z_{x,t}$ (rodents caught per 100 trap nights), for season x in year t is assumed to be an unbiased measure proportional to the unobserved rodent abundance, where $n_{x,t} = \ln(z_{x,t} + \tau_{1,t})$ and the transformation parameter τ is Beta(β_1, β_2) distributed (see Supplementary Information). The parameter τ represents low, random, abundances when no animals were caught. All statistics reported here are the mean results over at least 10^3 random series of τ. By H_t, U_t, and K_t we respectively denote the ground snow hardness, the relative humidity in April and the percentage snow cover in July.

We then fit a statistical population model describing the seasonal rodent abundance fluctuations:

$$z_{s,t} = \exp\left(\alpha_0 + \alpha_1 n_{a,t-1} + \alpha_2 h_t + \alpha_3 u_t + f_1(n_{s,t-1}) + \varepsilon_{s,t}\right) \quad (1)$$

$$z_{a,t} = \exp\left(\alpha_4 + \alpha_5 n_{a,t-1} + \alpha_6 k_t + \alpha_7 u_t + f_2(n_{s,t}) + \varepsilon_{a,t}\right) \quad (2)$$

Here $h_t = \ln(H_t + c_1)$, $k_t = \ln(K_t + c_2)$, $u_t = \ln(U_t)$, c_1 and c_2 are transformation constants, $f_x(y)$ represent nonlinear effects estimated from penalized regression splines (see Methods and Supplementary Information), $\varepsilon_{x,t}$ are quasi-Poissonian noise terms to allow for overdispersion, and $\alpha_0, \ldots, \alpha_7$ are estimated regression coefficients.

We find this to be an adequate model for the Finse rodents: it explains about 90% of the observed variance in catch rates when trained on the 1970–1997 data, correctly identifies all peak years with no false positives between 1970 and 1997 when doing one-step-ahead predictions, and correctly predicts an absence of peaks between 1998 and 2007, owing to its correctly predicting low spring abundances ($r = 0.70$, $n = 10$, $P < 0.05$). No significant serial autocorrelations were observed in the residuals of the seasonal models. Even when trained only on 1970–1990 data, this population model captures the peak years 1991 and 1994, as well as the absence of peaks thereafter.

Autumn catches are not well predicted in the low-abundance period, 1995–2008, as they are much more weakly coupled to spring catches during this period (spring–autumn 1970–1995: $r = 0.81$, $n = 26$, $P < 0.01$; 1996–2007: $r = 0.28$, $n = 12$, $P > 0.2$), but see Supplementary Information.

Model validation was performed by fitting on parts of the data set and predicting the remaining part, both by one-step-ahead predictions and multi-year simulations. The models exhibit mostly stationary dynamical behaviour over time. Model coefficients and diagnostics are given in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Stenseth, N. C. The long-term study of voles, mice and lemmings: homage to Robert Collett. *Trends Ecol. Evol.* **10**, 512 (1995).
2. Stenseth, N. C. & Ims, R. A. (eds) *The Biology of Lemmings*. (Academic, 1993).
3. Ims, R. A. & Fuglei, E. Trophic interaction cycles in tundra ecosystems and the impact of climate change. *Bioscience* **55**, 311–322 (2005).
4. Holt, R. D. Theoretical perspectives on resource pulses. *Ecology* **89**, 671–681 (2008).
5. Coulson, T. *et al.* Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* **292**, 1499–1500 (2001).
6. Framstad, E., Stenseth, N. C., Bjørnstad, O. N. & Falck, W. Limit cycles in Norwegian lemmings: tensions between phase-dependence and density-dependence. *Proc. R. Soc. Lond. B* **264**, 31–38 (1997).
7. Hörnfeldt, B. Long-term decline in numbers of cyclic voles in boreal Sweden: analysis and presentation of hypotheses. *Oikos* **107**, 376–392 (2004).
8. Hörnfeldt, B., Hipkiss, T. & Eklund, U. Fading out of vole and predator cycles? *Proc. R. Soc. B* **272**, 2045–2049 (2005).
9. Ims, R. A., Henden, J. A. & Killengreen, S. T. Collapsing population cycles. *Trends Ecol. Evol.* **23**, 79–86 (2008).
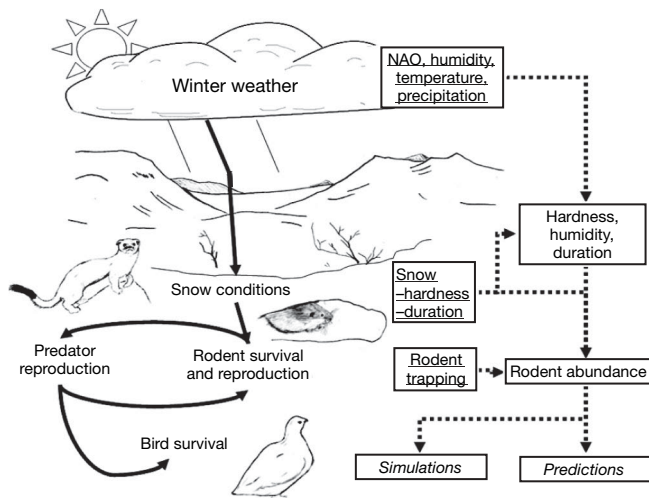
Figure 4 | Overview. Underlined text indicates available data, used in models (boxes) linking meteorological data, snow conditions and population dynamics. The resulting inferences (dotted lines) are consistent with processes (solid arrows) likely to be important for ecosystem dynamics.

10. Klemola, T., Pettersen, T. & Stenseth, N. C. Trophic interactions in population cycles of voles and lemmings: a modelling study and a review-synthesis. *Adv. Ecol. Res.* **33,** 75–160 (2003).

11. Steen, H. *et al.* Mortality of lemmings, *Lemmus lemmus,* at peak density in a mountainous area of Norway. *J. Zool.* **243,** 831–835 (1997).

12. Turchin, P. & Hanski, I. Contrasting alternative hypotheses about rodent cycles by translating them into parameterized models. *Ecol. Lett.* **4,** 267–276 (2001).

13. Strann, K.-B., Yoccoz, N. G. & Ims, R. A. Is the heart of Fennoscandian rodent cycle still beating? A 14-year study of small mammals and Tengmalm's owls in northern Norway. *Ecography* **25,** 81–87 (2002).

14. Bierman, S. N. *et al.* Changes over time in the spatiotemporal dynamics of cyclic populations of field voles (*Microtus agrestis* L.). *Am. Nat.* **167,** 583–590 (2006).

15. Aars, J. & Ims, R. A. Intrinsic and climatic determinants of population demography: The winter dynamics of tundra voles. *Ecology* **83,** 3449–3456 (2002).

16. Korslund, L. & Steen, H. Small rodent winter survival: snow conditions limit access to food resources. *J. Anim. Ecol.* **75,** 156–166 (2006).

17. Lindström, E. R. & Hörnfeldt, B. Vole cycles, snow depth, and fox predation. *Oikos* **70,** 156–160 (1994).

18. MacLean, S. F., Fitzgerald, B. M. & Pitelka, F. A. Population cycles in Arctic lemmings: winter reproduction and predation by weasels. *Arctic and Alpine Research* **6,** 1–12 (1974).

19. Marchand, P. J. *Life in the Cold: An Introduction to Winter Ecology* (University Press of New England, 1996).

20. Merritt, J. F. & Merritt, J. M. Population ecology and energy relationships of Clethrionomys gapperi in a Colorado subalpine forest. *J. Mamm.* **59,** 576–598 (1978).

21. Oksanen, L. & Oksanen, T. Long-term microtine dynamics in North Fennoscandian tundra – the vole cycle and the lemming chaos. *Ecography* **15,** 226–236 (1992).

22. Stenseth, N. C., Chan, K.-S., Framstad, E. & Tong, H. Phase- and density-dependent population dynamics in Norwegian lemmings: Interaction between deterministic and stochastic processes. *Proc. R. Soc. Lond.* **265,** 1957–1968 (1998).

23. Hanski, I. *et al.* Small-rodent dynamics and predation. *Ecology* **82,** 1505–1520 (2001).

24. Niklasson, B. *et al.* Diabetes and myocarditis in voles and lemmings at cyclic peak densities – induced by Ljungan virus? *Oecologia* **150,** 1–7 (2006).

25. Anderson, R. M. & May, R. M. Population biology of infectious diseases: Part I. *Nature* **280,** 361–367 (1979).

26. Gilg, O. *et al.* Functional and numerical responses of four lemming predators in high arctic Greenland. *Oikos* **113,** 193–216 (2006).

27. Henttonen, H., Oksanen, T., Jortikka, A. & Haukisalmi, V. How much do weasels shape microtine cycles in the northern Fennoscandian taiga? *Oikos* **50,** 353–365 (1987).

28. Lambin, X., Bretagnolle, V. & Yoccoz, N. G. Vole population cycles in northern and southern Europe: Is there a need for different explanations for single pattern? *J. Anim. Ecol.* **75,** 340–349 (2006).

29. Nordli, Ø., Lundstad, E. & Ogilvie, A. E. J. A late-winter to early-spring temperature reconstruction for southeastern Norway 1758 to 2006. *Ann. Glaciol.* **46,** 404–408 (2007).

30. Hanssen-Bauer, I., Førland, E. J., Haugen, J. E. & Tveito, O. E. Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Clim. Res.* **25,** 15–27 (2003).

METHODS

Statistical analyses. The strong seasonality of the system suggests that discrete time dynamics are applicable[31], and we assume that ease of trapping is unbiased through time, as trapping has taken place in the same survey programme in permanent plots of stable vegetation (see Supplementary Information). There is no significant between-year autocorrelation in the lemming spring densities, and only at year $t-2$ for the autumn densities ($r = -0.47$, $n = 38$, $P < 0.01$), where n denotes the number of data pairs.

We used generalized additive models (GAMs) with integrated smoothness estimation using penalized regression splines. In all GAMs, the nonlinear functions were constrained to have monotonic behaviour. Quasi-Poissonian error distributions were used to allow for overdispersion. To avoid noise and potential bias from the proxy humidity data, the GAMs were fitted on only the 1970–1997 data. None of the partly continuous environmental time series (temperature anomalies, snow cover, NAO and humidity) exhibited any significant between-year autocorrelation over the period 1969–2007, and the population model residuals were free from temporal autocorrelation.

The significance of the hardness and humidity covariates and their dynamical effects are robust over a variety of model formulations and approaches, including Bayesian state-space modelling. Temperature and NAO measures were also tried directly as covariates, but on the whole were found to perform worse and less robustly than the snow parameters, as would be expected if these were closer to the actual mechanisms.

All effects are, unless otherwise noted in the text, significant at the 5% level or less. Parameter tables and model diagnostics can be found in the Supplementary Information. Wavelet analyses using a Morlet wavelet and Beta surrogate significance test[32] were performed to assess changes in periodicity and coherence[33]. Analyses were performed using the software R (http://www.r-project.org).

Time series data on rodents and birds. The rodent data are 38-yr-long, seasonal trapping series from Finse, which is situated in the Hardangervidda massif of southern Norway (Fig. 1a) between 1200–1350 metres above sea level in the low- and mid-alpine zones[34]. Small mammals were monitored through trapping in two 1×1-ha^2 grids with 10×10 trap stations at 10-m intervals[6,35]. There were two periods of 4–6 days, the first in June–July (phenologically spring) and the second in August–September (phenologically autumn). All traps were checked daily. Lemmings were most frequently caught ($z_{mean} = 1.8$), but *Microtus oeconomus* ($z_{mean} = 0.45$), *Microtus agrestis* ($z_{mean} = 0.09$), *Sorex* ssp. ($z_{mean} = 0.10$), *Myodes glareolus* ($z_{mean} = 0.06$) and *Myodes rufocanus* ($z_{mean} = 0.02$) were also common. As preliminary analysis suggested that the Soricidae may respond somewhat atypically to the rest of the rodent group, probably owing to diet and metabolic rate differences as well as often being secondary prey relative to the rodents, they were not pooled with the non-lemming rodents and, hence, were not included in further analysis.

We also used the mean number of occupied passerine bird territories per square kilometre along three nearby transects and the number of occupied wader territories per square kilometre in the Finsefetene mudflats. These were gathered by repeated surveys[36,37] around the beginning of July 1967 until 1984 and 1985 for the waders and passerines, respectively. The data were pooled across species and transects.

The rodent trapping grids, the bird transects and mudflats and the Finse meteorological station all fit within an approximately 5×5-km^2 area to the south and east of the Finse railway station ($60.602°$ N, $7.504°$ E).

Hunter-reported catches of ptarmigan (*Lagopus muta*) and willow grouse (*Lagopus lagopus*) were obtained from the Norwegian Bureau of Statistics (http://www.ssb.no). As there have been significant changes in reporting procedures and hunting behaviour that may induce low-frequency trends in the data, we use the logarithmic rate of change from one year to the next as the most reliable data, as well as a GAM-detrended version (these data transformations correlate closely ($r \approx 0.70$) and give qualitatively very similar results).

None of the logarithmic rates of change for the passerines, waders or ptarmigan and willow grouse showed significant temporal autocorrelation.

Climate and snow conditions. Here we use the extended winter NAO index of Hurrell[38,39] (December in year $t-1$ until March in year t), based on the difference between normalized sea-level pressure in Lisbon, Portugal and Stykkisholmur/Reykjavik, Iceland, together with meteorological records[40]. Also, snow data were sampled as part of winter-ecology courses held at Finse in March–April on 15 occasions during 1970–2008. These were organized by three of the authors (E.Ø., I.M. and T.S.), providing first-hand information on the average hardness, measured by penetrometers as the pressure (in kilograms per square centimetre) needed to make an indentation in the snow layer closest

to the ground[19]. Field estimates of the percentage of ground that is snow-covered around the 10th of July were also made every year during 1970–2000 by one of the authors (E.Ø.).

Using daily temperature maxima ($T_{i,t}^{\mathrm{max,\,d}}$) and minima ($T_{i,t}^{\mathrm{min,\,d}}$) we sum the constants θ_1 and θ_2, which represent the daily contributions to snow hardness (that is, the opposite of subnivean space formation), over the days i to find the temperature fluctuation impact $\bar{T}_t$ on snow hardness year t:

$$\bar{T}_t = \begin{cases} \sum_i T_{i,t}^{\mathrm{max,\,d}} - T_{i,t}^{\mathrm{min,\,d}} & \text{if } T_{i,t}^{\mathrm{min,\,d}} < -3,\ T_{i,t}^{\mathrm{max,\,d}} > 0 \\ \sum_i \theta_1 & \text{if } T_{i,t}^{\mathrm{min,\,d}} > -3,\ T_{i,t}^{\mathrm{max,\,d}} > 0 \\ \sum_i \theta_2 & \text{if } T_{i,t}^{\mathrm{max,\,d}} < 0 \end{cases} \quad (3)$$

Together with monthly averages of temperature maxima ($T_{j,t}^{\mathrm{max,\,m}}$), medians ($T_{j,t}^{\mathrm{med,\,m}}$) and minima ($T_{j,t}^{\mathrm{min,\,m}}$) for month j, year t, this model was found to explain about 68% of the observed variance in mean measured hardness:

$$H_t = \exp\left(\theta_0 + f_3(\bar{T}_t) + f_4(T_{2,t}^{\mathrm{med,m}} + T_{3,t}^{\mathrm{max,m}} + T_{4,t}^{\mathrm{min,m}} + T_{5,t}^{\mathrm{min,m}} + T_{10,t-1}^{\mathrm{max,m}}) + \varepsilon_t\right) (4)$$

The NAO, the precipitation in April in millimetres ($P_{4,t}$) together with monthly temperatures $\bar{T}'_t = T_{3,t}^{\mathrm{med,m}} + c_3 T_{4,t}^{\mathrm{med,m}} + T_{5,t}^{\mathrm{med,m}}$ explain about 74% of the observed variance in median relative humidity in April (U_t):

$$U_t = \frac{100}{1 + \exp\left(-1(v_0 + v_1 \mathrm{NAO}_t + v_2 \bar{T}'_t + v_3 P_{4,t} + v_4 P_{4,t} \bar{T}'_t) + f_5(T_{4,t}^{\mathrm{max,m}} - T_{4,t}^{\mathrm{min,m}}) + \varepsilon_t\right)} \quad (5)$$

The effects of NAO and temperature explain about 87% of the observed variance in July snow cover:

$$K'_t = \frac{100}{1 + \exp\left(-1(\kappa_0 + f_6(\mathrm{NAO}_t) + f_7(T_{5,t}^{\mathrm{max,m}}) + f_8(T_{6,t}^{\mathrm{max,m}}) + f_9(T_{10,t-1}^{\mathrm{max,m}})) + \varepsilon_t\right)} \quad (6)$$

Above, θ_0, κ_0 and $v_0, \ldots, v_4$ are estimated regression parameters, θ_1, θ_2 and c_3 are weighting constants. All parameters can be found in the Supplementary Information.

Simulations. As climatic fluctuations normally will prevent equilibrium states from being dominant, the transient dynamics are of ecological interest[4,31,41,42]. Hence, the dynamics captured by our population models were assessed through stochastic simulations (that is, using only the previous year's predicted population values when predicting the next, and adding random errors from the estimates distribution of the residuals). These were simulated over 100 yr for each of 10^3 different climate regimes generated by skewing their empirical probability distributions towards higher or lower values but not going beyond the observed range. The number of years between spring and/or autumn catch rates exceeding one lemming per 100 trap nights was adopted as a practical definition of cycle length.

31. Geritz, S. A. H. & Kisdi, E. On the mechanistic underpinning of discrete-time population models with complex dynamics. *J. Theor. Biol.* **228**, 261–269 (2004).
32. Rouyer, T., Fromentin, J.-M., Stenseth, N. C. & Cazelles, B. Analysing multiple time series and extending significance testing in wavelet analysis. *Mar. Ecol. Prog. Ser.* **359**, 11–23 (2008).
33. Cazelles, B. *et al.* Wavelet analysis of ecological time series. *Oecologia* **156**, 287–304 (2008).
34. Østbye, E. *et al.* Hardangervidda. *Ecol. Bull.* **20**, 225–264 (1975).
35. Stenseth, N. C. Population cycles in voles and lemmings: density dependence and phase dependence in a stochastic world. *Oikos* **87**, 427–461 (1999).
36. Østbye, E. *et al.* Structure and dynamics of some high mountain bird communities of South Norway: a 19-year study of passerines. *Ornis Norvegica* **25**, 19–48 (2002).
37. Østbye, E. *et al.* Structure and dynamics of a high mountain wetland bird community in southern Norway: An 18-year study of waders and gulls. *Ornis Norvegica* **30**, 4–20 (2007).
38. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
39. Stenseth, N. C. *et al.* Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proc. R. Soc. Lond. B* **270**, 2087–2096 (2003).
40. Norwegian Meteorological Institute. ⟨http://www.met.no⟩ (2008).
41. Schmidt, K. A. & Ostfeld, R. S. Numerical and behavioral effects within a pulse-driven system: consequences for shared prey. *Ecology* **89**, 635–646 (2008).
42. Dennis, B. *et al.* Can noise induce chaos? *Oikos* **102**, 329–339 (2003).

LETTERS

Genes mirror geography within Europe

John Novembre[1,2], Toby Johnson[4,5,6], Katarzyna Bryc[7], Zoltán Kutalik[4,6], Adam R. Boyko[7], Adam Auton[7], Amit Indap[7], Karen S. King[8], Sven Bergmann[4,6], Matthew R. Nelson[8], Matthew Stephens[2,3] & Carlos D. Bustamante[7]

Understanding the genetic structure of human populations is of fundamental interest to medical, forensic and anthropological sciences. Advances in high-throughput genotyping technology have markedly improved our understanding of global patterns of human genetic variation and suggest the potential to use large samples to uncover variation among closely spaced populations[1–5]. Here we characterize genetic variation in a sample of 3,000 European individuals genotyped at over half a million variable DNA sites in the human genome. Despite low average levels of genetic differentiation among Europeans, we find a close correspondence between genetic and geographic distances; indeed, a geographical map of Europe arises naturally as an efficient two-dimensional summary of genetic variation in Europeans. The results emphasize that when mapping the genetic basis of a disease phenotype, spurious associations can arise if genetic structure is not properly accounted for. In addition, the results are relevant to the prospects of genetic ancestry testing[6]; an individual's DNA can be used to infer their geographic origin with surprising accuracy—often to within a few hundred kilometres.

Recent studies suggest that by combining high-throughput genotyping technologies with dense geographic samples one can shed light on unanswered questions regarding human population structure[1–5]. For instance, it is not clear to what extent populations within continental regions exist as discrete genetic clusters versus as a genetic continuum, nor how precisely one can assign an individual to a geographic location on the basis of their genetic information alone.

To investigate these questions, we surveyed genetic variation in a sample of 3,192 European individuals collected and genotyped as part of the larger Population Reference Sample (POPRES) project[7]. Individuals were genotyped at 500,568 loci using the Affymetrix 500K single nucleotide polymorphism (SNP) chip. When available, we used the country of origin of each individual's grandparents to determine the geographic location that best represents each individual's ancestry, otherwise we used the self-reported country of birth (see Methods and Supplementary Tables 1 and 2). After removing SNPs with low-quality scores, we applied various stringency criteria to avoid sampling individuals from outside of Europe, to create more even sample sizes across Europe, to exclude individuals with grandparental ancestry from more than location, and to avoid potential complications of SNPs in high linkage disequilibrium (see Methods and Supplementary Table 3). Although our main result holds even when we relax nearly all of these stringency criteria, we focus our analyses on genotype data from 197,146 loci in 1,387 individuals (Supplementary Table 2), for whom we have high confidence of individual origins.

We used principal components analysis (PCA; ref. 8) to produce a two-dimensional visual summary of the observed genetic variation.

The resulting figure bears a notable resemblance to a geographic map of Europe (Fig. 1a). Individuals from the same geographic region cluster together and major populations are distinguishable. Geographically adjacent populations typically abut each other, and recognizable geographical features of Europe such as the Iberian peninsula, the Italian peninsula, southeastern Europe, Cyprus and Turkey are apparent. The data reveal structure even among French-, German- and Italian-speaking groups within Switzerland (Fig. 1b), and between Ireland and the United Kingdom (Fig. 1a, IE and GB). Within some countries individuals are strongly differentiated along the principal component (PC) axes, suggesting that in some cases the resolution of the genetic data may exceed that of the available geographic information.

When we quantitatively compare the geographic position of countries with their PC-based genetic positions, we observe few prominent differences between the two (Supplementary Fig. 1), and those that exist can be explained either by small sample sizes (for example, Slovakia (SK)) or by the coarseness of our geographic data (a problem for large countries, for example, Russia (RU)); see Supplementary Information for more detail. Our method also identifies a few individuals who exhibit large differences between their genetic and geographic positions (Supplementary Fig. 2). These individuals may have mis-specified ancestral origins or be recent migrants. In addition, although the sample used here is unlikely to include many members of smaller genetically isolated populations that exist within countries (for example, Basque residing in Spain or France, Orcadians in Scotland, or individuals of Jewish ancestry), in rare cases outlying individuals could reflect membership of such groups. For example, a small set of Italian individuals cluster 'southwest' of the main Italian cluster and one might speculate they are individuals of insular Italian origin (for example, Sardinia or Sicily).

The overall geographic pattern in Fig. 1a fits the theoretical expectation for models in which genetic similarity decays with distance in a two-dimensional habitat, as opposed to expectations for models involving discrete well-differentiated populations. Indeed, in these data genetic correlation between pairs of individuals tends to decay with distance (Fig. 1c). For spatially structured data, theory predicts the top two principal components (PCs 1 and 2) to be correlated with perpendicular geographic axes[9], which is what we observe ($r^2 = 0.71$ for PC1 versus latitude; $r^2 = 0.72$ for PC2 versus longitude; after rotation, $r^2 = 0.77$ for 'north–south' in PC-space versus latitude, and $r^2 = 0.78$ for 'east–west' in PC-space versus longitude). In contrast, when there are K discrete populations sampled, one expects discrete clusters to be separated out along $K - 1$ of the top PCs[8]. In our analysis, neither the first two PCs, nor subsequent PCs, separate clusters as one would expect for a set of discrete, well-differentiated populations (see ref. 8 for examples).

[1]Department of Ecology and Evolutionary Biology, Interdepartmental Program in Bioinformatics, University of California–Los Angeles, Los Angeles, California 90095, USA. [2]Department of Human Genetics, [3]Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA. [4]Department of Medical Genetics, [5]University Institute for Social and Preventative Medecine, Centre Hospitalier Universitaire Vaudois (CHUV), University of Lausanne, Rue de Bugnon 27 - DGM 328, CH-1005 Lausanne, Switzerland. [6]Swiss Institute of Bioinformatics, Central Administration, Quartier Sorge - Batiment Genopode, 1015 Lausanne, Switzerland. [7]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA. [8]GlaxoSmithKline, Research Triangle Park, North Carolina 27709, USA.

The direction of the PC1 axis and its relative strength may reflect a special role for this geographic axis in the demographic history of Europeans (as first suggested in ref. 10). PC1 aligns north-northwest/south-southeast (NNW/SSE, −16 degrees) and accounts for approximately twice the amount of variation as PC2 (0.30% versus 0.15%, first eigenvalue = 4.09, second eigenvalue = 2.04). However, caution is required because the direction and relative strength of the PC axes are affected by factors such as the spatial distribution of samples (results not shown, also see ref. 9). More robust evidence for the importance of a roughly NNW/SSE axis in Europe is that, in these same data, haplotype diversity decreases from south to north (A.A. *et al.*, submitted). As the fine-scale spatial structure evident in Fig. 1 suggests, European DNA samples can be very informative about the geographical origins of their donors. Using a multiple-regression-based assignment approach, one can place 50% of

individuals within 310 km of their reported origin and 90% within 700 km of their origin (Fig. 2 and Supplementary Table 4, results based on populations with $n > 6$). Across all populations, 50% of individuals are placed within 540 km of their reported origin, and 90% of individuals within 840 km (Supplementary Fig. 3 and Supplementary Table 4). These numbers exclude individuals who reported mixed grandparental ancestry, who are typically assigned to locations between those expected from their grandparental origins (results not shown). Note that distances of assignments from reported origin may be reduced if finer-scale information on origin were available for each individual.

Population structure poses a well-recognized challenge for disease-association studies (for example, refs 11–13). The results obtained here reinforce that the geographic distribution of a sample is important to consider when evaluating genome-wide association studies
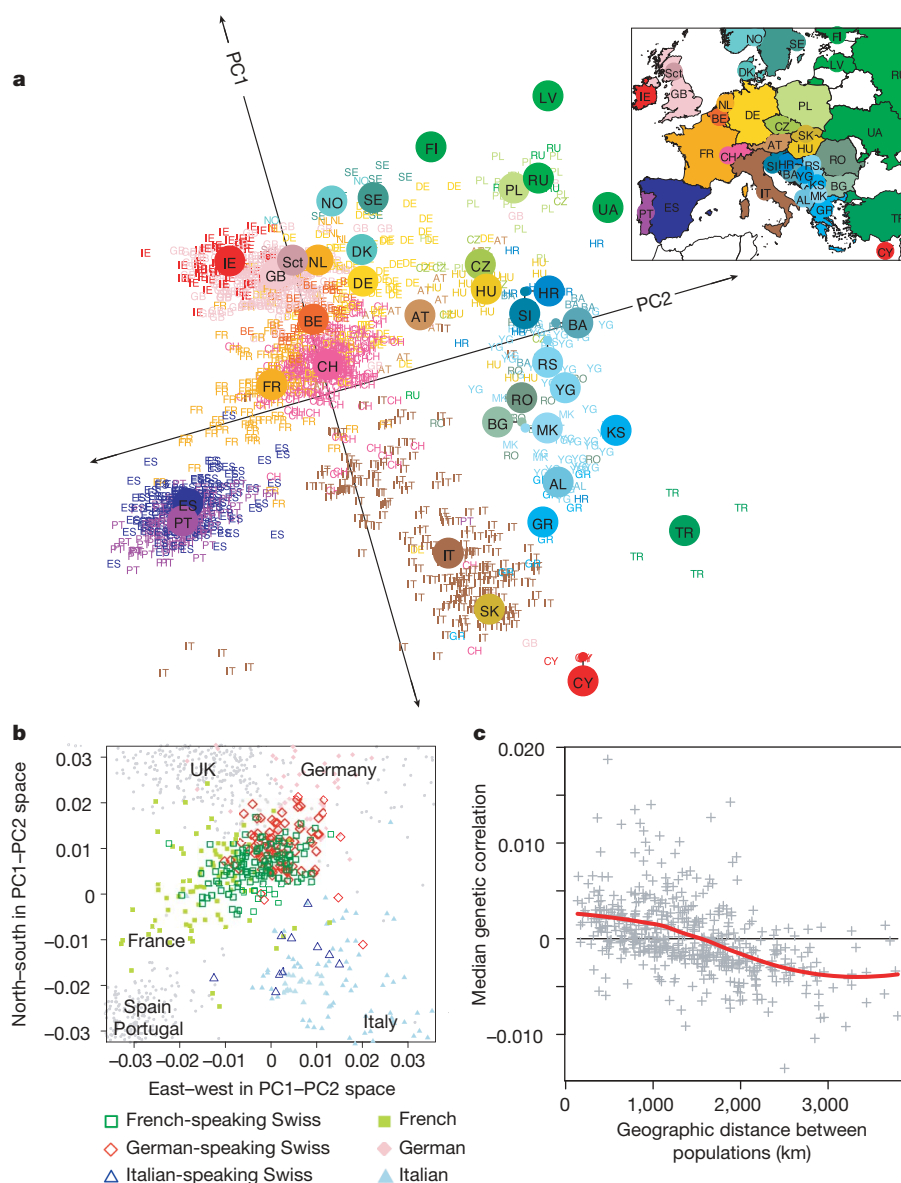

Figure 1 | Population structure within Europe. a, A statistical summary of genetic data from 1,387 Europeans based on principal component axis one (PC1) and axis two (PC2). Small coloured labels represent individuals and large coloured points represent median PC1 and PC2 values for each country. The inset map provides a key to the labels. The PC axes are rotated to emphasize the similarity to the geographic map of Europe. AL, Albania; AT, Austria; BA, Bosnia-Herzegovina; BE, Belgium; BG, Bulgaria; CH, Switzerland; CY, Cyprus; CZ, Czech Republic; DE, Germany; DK, Denmark; ES, Spain; FI, Finland; FR, France; GB, United Kingdom; GR, Greece; HR,

Croatia; HU, Hungary; IE, Ireland; IT, Italy; KS, Kosovo; LV, Latvia; MK, Macedonia; NO, Norway; NL, Netherlands; PL, Poland; PT, Portugal; RO, Romania; RS, Serbia and Montenegro; RU, Russia, Sct, Scotland; SE, Sweden; SI, Slovenia; SK, Slovakia; TR, Turkey; UA, Ukraine; YG, Yugoslavia. **b**, A magnification of the area around Switzerland from **a** showing differentiation within Switzerland by language. **c**, Genetic similarity versus geographic distance. Median genetic correlation between pairs of individuals as a function of geographic distance between their respective populations.

among Europeans (for example, refs 3–5, 11). A crucial part is also played by spatial variation in phenotype. To examine this, we simulated genome-wide association data for quantitative trait phenotypes with varying degrees of linear latitudinal or longitudinal trends (Supplementary Fig. 4). Even for phenotypes modestly correlated with geography (for example, ≥5% of variance explained by latitude or longitude) the uncorrected P-value distribution shows a clear excess of small values, suggesting that population structure correction may be important even in seemingly closely related populations such as Europeans. Note that many factors, including sample size and distribution of sampling locations, will influence the effects of stratification on P-value distributions, and so these results should be considered only as illustrative of the settings in which stratification could become a problem in European samples.

In all our simulations, use of a PC-based correction[12,14] adequately controlled for P-value inflation (Supplementary Fig. 4). The success of PCA-based correction is not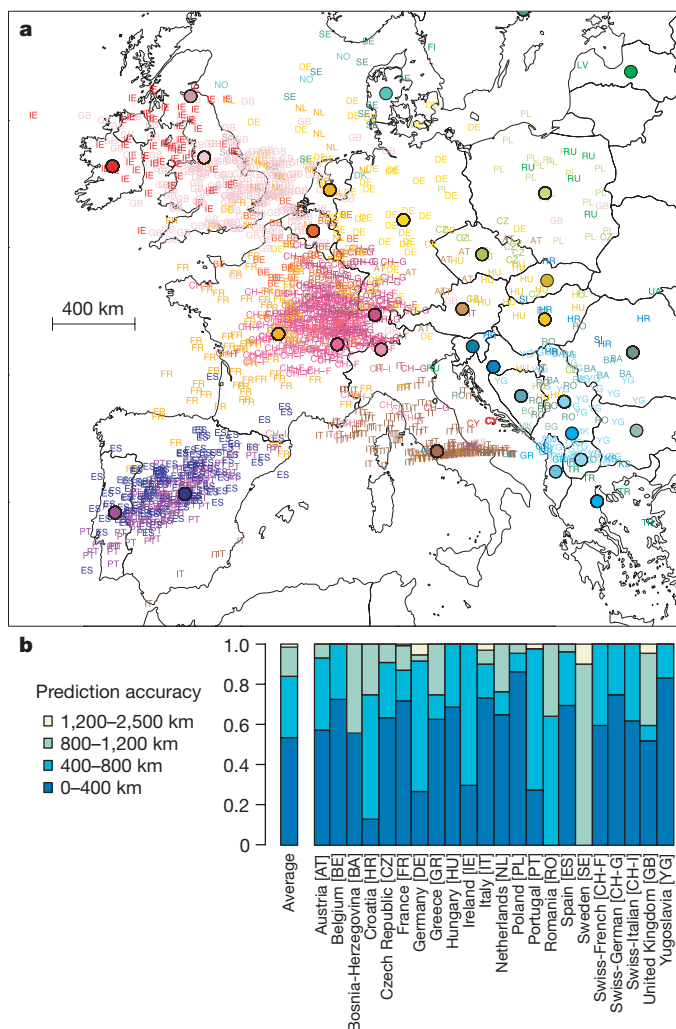 unexpected here, because the PCs are excellent predictors of latitude and longitude, and we used only linear functions of latitude and longitude to determine the means of our simulated phenotypes. For real phenotypes, higher order functions of PC1 and PC2 and/or additional PCs might be necessary to correct for more complex spatial variation in phenotype. We speculate that at the geographic scale of many association studies carried out so far, many phenotypes are relatively uncorrelated with geography, and that this may explain why in many cases PC-based correction has had little impact in practice[3,13]. For phenotypes that are more strongly spatially structured within a sample (for example, height[11,15,16]), spurious associations due to population stratification should be more of a concern.

Although broad correlations between PCs and geography have been observed previously[3–5,17,18] only the large number of loci and dense geographic sampling of individuals used here reveal the clear map-like structure to European genetic variation. Because at any one SNP the average level of differentiation across Europe is small (average $F_{ST} = 0.004$ between geographic regions; F_{ST} is a measure of differentiation between populations that takes values of 0 when there is no differentiation and one when there is maximal differentiation[19]), it is the combined information across many loci and many individuals that reveals fine-scale population structure in this sample.

An important consideration in interpreting our analyses is that, as a result of ascertainment bias[20,21], current SNP genotyping platforms under-represent variation at low-frequency alleles. Low-frequency alleles tend to be the result of a recent mutation and are expected to geographically cluster around the location at which the mutation first arose; hence, they can be highly informative about the fine-scale population structure (for example, ref. 22). In addition, the PCA-based methods used here are based on genotypic patterns of variation and do not take advantage of signatures of population structure that are contained in patterns of haplotype variation[1,23–25]. Soon-to-be-available whole-genome re-sequencing will give us access to informative low-frequency alleles, and further statistical method development will allow us to leverage patterns of haplotype variation. The prospect of these developments suggests the geographic resolution presented here is only a lower bound on the performance possible in the near future. Thus, our results provide an important insight: the power to detect subtle population structure, and in turn the promise of genetic ancestry tests, may be more substantial than previously imagined.

METHODS SUMMARY

The sample of European individuals used here was assembled and genotyped as part of the larger POPRES project[7]. Genotyping was carried out using the Affymetrix GeneChip Human Mapping 500K Array Set. No significant differentiation was observed between individuals collected and/or genotyped at different times (analysis of variance, ANOVA, $P > 0.05$).

PCA was carried out using the smartpca program[8,12]. Before running PCA, we removed SNPs that showed evidence of high pairwise linkage disequilibrium as well as unique genomic regions (such as large polymorphic inversions) that might obscure genome-wide patterns of population structure. In addition, an initial PCA run was used to remove extreme genetic outliers.

When comparing the PC results to geography, we assigned each individual a location—typically the geographic centre of their corresponding population (Supplementary Table 3). The rotation of axes used in Fig. 1 is 16 degrees counterclockwise and was determined by finding the angle that maximizes the summed correlation of the median PC1 and PC2 values with the latitude and longitude of each country.

The new assignment method used here is based on independent linear models for latitude and longitude where each is predicted jointly by PC1 and PC2, including quadratic terms and an interaction term. To assess performance, we used leave-one-out cross-validation and adjusted for unequal sample sizes (for example, we weigh each population equally when computing the mean prediction accuracy).

For the genome-wide association simulations, we simulated each individual's phenotype as having a mean determined by his or her geographic position and then simulated Gaussian distributed residual variation to obtain a phenotype with a fixed proportion of variance explained by geographic position. To perform the association test with PC-based correction, we used multiple linear

Figure 2 | Performance of assignment method. a, Predicted locations for each of 1,387 individuals based on leave-one-out cross validation and the continuous assignment method. Small coloured labels (for definitions, see Fig. 1 legend, except here CH-I, CH-F, and CH-G denote Swiss individuals who speak Italian, French, or German respectively) represent individual assignments. Coloured points denote the locations used to train the assignment method. **b,** Distribution of prediction accuracy by country. Distances are measured between the population assigned by the discrete assignment method and the geographic origin of the individual. The average is taken of the proportions across populations and each population is given equal weight. The panel shows results for populations with greater than six individuals; performance decreases for populations with smaller sample sizes (Supplementary Fig. 3).

regression with PC1 and PC2 as covariates, as implemented in the program eigenstrat[8,12].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
2. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
4. Tian, C. *et al.* Analysis and application of European genetic substructure using 300K SNP information. *PLoS Genet.* **4**, e4 (2008).
5. Price, A. L. *et al.* Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet.* **4**, e236 (2008).
6. Shriver, M. D. & Kittles, R. A. Genetic ancestry and the search for personalized genetic histories. *Nature Rev. Genet.* **5**, 611–618 (2004).
7. Nelson, M. R. *et al.* The Population Reference Sample (POPRES): a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* (in the press).
8. Patterson, N., Price, A. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
9. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nature Genet.* **40**, 646–649 (2008).
10. Menozzi, P., Piazza, A. & Cavalli-Sforza, L. Synthetic maps of human gene frequencies in Europeans. *Science* **201**, 786–792 (1978).
11. Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature Genet.* **37**, 868–872 (2005).
12. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38**, 904–909 (2006).
13. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
14. Zhu, X., Zhang, S., Zhao, H. & Cooper, R. S. Association mapping, using a mixture model for complex traits. *Genet. Epidemiol.* **23**, 181–196 (2002).
15. Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genet.* **40**, 575–583 (2008).
16. Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genet.* **40**, 584–591 (2008).
17. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* 292 (Princeton Univ. Press, 1994).
18. Bauchet, M. *et al.* Measuring European population stratification with microarray genotype data. *Am. J. Hum. Genet.* **80**, 948–956 (2007).
19. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
20. Eberle, M. A. & Kruglyak, L. An analysis of strategies for discovery of single nucleotide polymorphisms. *Genet. Epidemiol.* **19**, S29–S35 (2000).
21. Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H. & Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* **15**, 1496–1502 (2005).
22. Slatkin, M. Rare alleles as indicators of gene flow. *Evolution* **39**, 53–65 (1985).
23. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
24. Tang, H., Coram, M., Wang, P., Zhu, X. & Risch, N. Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* **79**, 1–12 (2006).
25. Hellenthal, G., Auton, A. & Falush, D. Inferring human colonization history using a copying model. *PLoS Genet.* **4**, e1000078 (2008).

METHODS

Sample collection and genotyping. The samples were assembled and genotyped as part of the larger POPRES project currently consisting of ~6,000 individuals from worldwide populations[7]. The subsample of European individuals used here is derived from two independent collections: the London Life Sciences Population (LOLIPOP) study[26], which consists mainly of European individuals sampled in London, and (2) the CoLaus study[27], which represents a broad set of European individuals sampled from Lausanne, Switzerland. The combined sample contains individuals with origins from across Europe (Supplementary Table 2), although origins from eastern Europe are generally less well represented (for example, Finland, Latvia, Ukraine, Slovakia and Slovenia) and some countries are not sampled at all (for example, Belarus, Estonia, Lithuania and Moldova).

Genotyping was carried out using the Affymetrix GeneChip Human Mapping 500K Array Set according to published protocol. We observe no significant differentiation in the PCA between individuals collected and/or genotyped at different times (ANOVA, $P > 0.05$). A thorough description of the collections, data processing methods and public data release is presented in ref. 7.

To prepare the sample analysed here, we used the demographic data available for each individual to create a 'geographic origin' that represents a single location from which the individual's very recent ancestry is derived. Where possible, we based the geographic origin on the observed country data for grandparents. We used a 'strict consensus' approach: if all observed grandparents originated from a single country, we used that country as the origin. If an individual's observed grandparents originated from different countries, we excluded the individual. Where grandparental data were unavailable, we used the individual's country of birth.

We excluded individuals whose putative geographic origin was from outside of Europe (for example, Europeans from USA, China, Mozambique, Ivory Coast, and so on), individuals who were putatively related (using the same approach as in ref. 7), and individuals found to be outliers in a preliminary PCA run (for more detail, see the section on PCA below). Because of the large number of Swiss individuals available and the availability of language information for most of these individuals, for some analyses, we divided Swiss individuals into three ancestry labels (Swiss-French, Swiss-German and Swiss-Italian) on the basis of their reported primary language. Finally, we chose to include only a random sample of 200 individuals from the United Kingdom and 125 Swiss-French to obtain more even sample sizes across Europe. Supplementary Table 2 provides more detail on how the sample numbers changed with each step in the sample preparation, and Supplementary Table 1 summarizes the number of grandparents observed for the 1,387 individuals used in the final sample.

Geographic locations associated with each country were assigned using the central point of the geographic area of the country (Supplementary Table 3). Three exceptions are the Russian Federation, Sweden and Norway, where the geographic locations were assigned to the location of the capitals of these countries (because central points were assumed to not be as reflective of the probable origins of these individuals). Within Switzerland, we represent the Swiss-French with the geographical coordinates of Geneva, the Swiss-German with those of Zurich, and Swiss-Italian with those of Lugano. Distances between points are always calculated as great circle distances.

For estimating F_{ST}[19] and for assessing the performance of assignment, we combined individuals into geographic groupings with larger and more comparable sample sizes than the original ancestral origins. These groupings do not reflect discrete structure in the data, rather the practical need to create geographical groupings with reasonable sample sizes. The strategy was to create a 3×3 grid of regions across Europe, with a tenth region for far southeastern Europe (Supplementary Table 3).

Principal components analysis. To conduct PCA, we used the smartpca software[8,12]. In a preliminary phase of the study, we ran smartpca using default settings and five outlier detection iterations, which resulted in the identification and exclusion of 34 individuals that were greater than six standard deviations from the mean PC position on at least one of the top ten eigenvectors. For our final run, we use the default settings without any outlier removal.

To avoid artefacts due to patterns of linkage disequilibrium[3], we filtered autosomal SNPs using two approaches simultaneously. First, before running PCA we used the PLINK[28] software to exclude SNPs with pairwise genotypic r^2 greater than 80% within sliding windows of 50 SNPs (with a 5-SNP increment between windows). Second, we took an iterative approach by running an initial PCA and removing chromosomal regions that showed evidence of reflecting regions of exceptional long-range linkage disequilibrium rather than genome-wide patterns of structure. These regions are detectable by plotting the correlation between individual PC scores and genotypes against the genome and identifying sharp, concentrated peaks in correlation (alternatively, we could have plotted the magnitude of elements of the SNP-based eigenvectors from the PCA,

but here we used the correlation-based approach because much of this work was done before the release of recent versions of smartpca that provide the SNP eigenvectors). SNPs falling within a 4 megabase region of a peak were excluded from the final PCA. Initially, peaks were defined by taking the top 0.01% of SNPs correlating with a PC for each of the top 6 PCs of the preliminary analysis. In this initial analysis PCs 1 and 2 did not appear to be artefacts of long-range linkage disequilibrium, but we still removed regions around the top PC-correlated SNPs. This approach is conservative (in the sense that we potentially remove more SNPs than necessary and hence might hinder ourselves from detecting subtle patterns). The procedure removed SNPs in regions such as the lactase region (2q21), the MHC region and the inversion regions 8p23 and 17q21.31, amongst others. The final number of SNPs used for PCA was 197,146 SNPs. The patterns of structure observed in PCs 1 and 2 were robust to further removal of chromosomal regions correlated with the PCs, suggesting the observed patterns are representative of genome-wide differentiation.

The inter-individual genetic correlations used in Fig. 1c were the same as those used for the PCA analysis and were obtained using the formula of ref. 8 as computed by smartpca.

The angle used to create the rotated PC1–PC2 coordinate system that is used in Fig. 1 was obtained by maximizing θ in the objective function:

$$f(\theta) = \mathrm{Cor}(g(\theta, \mathbf{v}_1, \mathbf{v}_2), \mathbf{Long}) + \mathrm{Cor}(h(\theta, \mathbf{v}_1, \mathbf{v}_2), \mathbf{Lat})$$

where $g(\theta, \mathbf{v}_1, \mathbf{v}_2)$ and $h(\theta, \mathbf{v}_1, \mathbf{v}_2)$ are functions that return coordinates of $\mathbf{v}_1$ (the PC1 eigenvector) and $\mathbf{v}_2$ (the PC2 eigenvector) after rotation about the point $(0,0)$ in PC1–PC2 space by the angle θ. **Lat** and **Long** are vectors of the latitude and longitude of each individual, and $\mathrm{Cor}(\cdot, \cdot)$ is the correlation function. The resulting optimal value of θ was found to be -16 degrees.

Spatial assignment. We assigned each individual to a specific geographic location by fitting independent linear models for latitude and longitude as predicted jointly by PC1 and PC2. We used the rotated PC1 and PC2 scores because these more strongly correlate with latitude and longitude (see main text). Specifically, we use the linear models:

$$\mathbf{x} = \beta_{x1}\mathbf{u}_1 + \beta_{x2}\mathbf{u}_2 + \beta_{x11}\mathbf{u}^2_1 + \beta_{x22}\mathbf{u}^2_2 + \beta_{x12}\mathbf{u}_1\mathbf{u}_2 + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \beta_{y1}\mathbf{u}_1 + \beta_{y2}\mathbf{u}_2 + \beta_{y11}\mathbf{u}^2_1 + \beta_{y22}\mathbf{u}^2_2 + \beta_{y12}\mathbf{u}_1\mathbf{u}_2 + \boldsymbol{\varepsilon}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors containing the longitude and latitude, respectively, of each individual, $\mathbf{u}_1$ and $\mathbf{u}_2$ are vectors containing the rotated PC1 and PC2 scores, respectively, for each individual (that is, $\mathbf{u}_1 = g(\theta, \mathbf{v}_1, \mathbf{v}_2)$, $\mathbf{u}_2 = h(\theta, \mathbf{v}_1, \mathbf{v}_2)$, where $\theta = -16$ degrees), β coefficients are regression coefficients, and $\boldsymbol{\varepsilon}$ represents residual error.

To perform assignment, we first estimated the β coefficients by means of least-squares regression with a training set of individuals with known locations and then used the estimated coefficients of the linear model to predict the latitude and longitude of a test individual on the basis of their PC1 and PC2 values (we call this a 'continuous assignment'). We also made a 'discrete assignment' by assigning individuals to the country for which the centre-point is closest to the latitude and longitude predicted by the continuous assignment method. In practice, the two methods produce roughly similar results (Supplementary Table 4). As a reference point for evaluating performance, the Supplementary Table also reports statistics for how a method would perform if all individuals were assigned to a central location within Europe (here taken to be Austria).

Simulation of genome-wide association study for a spatially structured quantitative trait. We simulated two types of traits: one with a latitudinal trend in the mean and the other with a longitudinal trend. For each type of trait, we simulated a range of different degrees to which the geographical axis (latitude or longitude) contributed to the overall variance in the trait. Specifically, we let $\mathbf{x}'$ and $\mathbf{y}'$ be normalized latitudinal and longitudinal variables, respectively (that is, $\mathbf{x}' = (\mathbf{x} - \bar{\mathbf{x}})/\sigma_x$ and $\mathbf{y}' = (\mathbf{y} - \bar{\mathbf{y}})/\sigma_y$, where $\mathbf{x}$ is a vector of each individual's longitude, $\mathbf{y}$ is likewise for latitude, $\bar{\mathbf{t}}$ is the mean value of the elements of $\mathbf{t}$, and σ_t is their standard deviation). We then simulated two phenotypes with the mean determined by $\mathbf{x}'$ or $\mathbf{y}'$: $\boldsymbol{\phi}_x = \mathbf{x}' + \boldsymbol{\varepsilon}_x$ and $\boldsymbol{\phi}_y = \mathbf{y}' + \boldsymbol{\varepsilon}_y$, where $\boldsymbol{\varepsilon}$ is a vector of random normal deviates with mean 0 and variance s^2. We let s^2 take values of (1, 4, 19, 99), so that the resulting variance in the traits are approximately (2, 5, 20, 100), and the proportion of variance explained is approximately (50, 20, 5, 1) per cent.

To perform the association test with PC-based correction, we used multiple linear regression with PC1 and PC2 as covariates as implemented in the software eigenstrat[12]. The Armitage χ^2 statistic was used to test the strength of the association. We also calculate an inflation statistic, by taking the ratio of the 50%

quantile of the observed Armitage χ^2 statistic with that expected under the null χ^2_1 distribution.

26. Kooner, J. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nature Genet.* **40**, 149–151 (2008).
27. Firmann, M. *et al.* The CoLaus study: A population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Dis.* **8**, 6 (2008).
28. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

CORRIGENDUM

The delayed rise of present-day mammals

Olaf R. P. Bininda-Emonds, Marcel Cardillo, Kate E. Jones,
Ross D. E. MacPhee, Robin M. D. Beck, Richard Grenyer,
Samantha A. Price, Rutger A. Vos, John L. Gittleman & Andy Purvis

Nature **446**, 507–512 (2007)

We have discovered a bug in the Perl script relDate v.2.2 that was used in part to date the nodes in the species-level mammalian supertree presented and analysed in our Article. The bug affected all but 80 of the 2,109 published dates, generally causing them to be slightly inflated, with the effect being stronger in more recent nodes. The absolute errors are mostly small (mean and median change of 1.32 and 0.70 million years, respectively), and a strong correlation between the two sets of dates exists ($r = 0.990$); however, 25 dates (all within Chiroptera) do change by more than 10 million years. Four of these dates are associated with the paraphyletic genus *Hipposideros*, whereas the remaining 21 cover most of Molossidae. The errors do not affect the results or overall conclusions of our paper qualitatively.

The Supplementary Information, including the tree files, has now been amended and can be accessed through the Supplementary Information link of the original Article. An additional file with a version of the amended Article can be accessed at http://www.uni-oldenburg.de/molekularesystematik/ under the 'Publikationen/ Publications' link.

CORRIGENDUM

STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling

Hiroki Ishikawa & Glen N. Barber

Nature **455**, 674–678 (2008)

We inadvertently failed to notice that STING protein is encoded by the same gene as the previously described plasma membrane tetra-spanner MPYS[1].

1. Jin, L. *et al.* MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex cass II and mediates transduction of apoptotic signals. *Mol. Cell. Biol.* **28**, 5014–5026 (2008).

CORRIGENDUM

A role for clonal inactivation in T cell tolerance to Mls-1ᵃ

Marcia A. Blackman, Hans-Gerhard Burgert, David L. Woodland,
Ed Palmer, John W. Kappler & Philippa Marrack

Nature **345**, 540–542 (1990)

In this Article, the name of Hans-Gerhard Burgert was incorrectly listed as Hans Gerhard-Burgert.

ADDENDUM

Genes mirror geography within Europe

John Novembre, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik,
Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King,
Sven Bergmann, Matthew R. Nelson, Matthew Stephens
& Carlos D. Bustamante

Nature **456**, 98–101 (2008)

A related manuscript arriving at broadly similar conclusions based on partially overlapping data has recently been published[1]. Specifically, 661 of the 3,192 samples from the POPRES collection[2] analysed in our paper were also analysed by Lao *et al.*[1].

1. Lao, O. *et al.* Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).
2. Nelson, M. R. *et al.* The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics. *Am. J. Hum. Genet.* **83**, 347–358 (2008).

CORRIGENDUM

Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains

Yusuke Sato, Azusa Yoshikawa, Atsushi Yamagata,
Hisatoshi Mimura, Masami Yamashita, Kayoko Ookata,
Osamu Nureki, Kazuhiro Iwai, Masayuki Komada & Shuya Fukai

Nature **455**, 358–362 (2008)

In this Fig. 3c of this Article, Asp 324 was incorrectly labelled as Glu 324.

LETTERS

Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval

Germán Sumbre[1]†, Akira Muto[2], Herwig Baier[2] & Mu-ming Poo[1]

The ability to process temporal information is fundamental to sensory perception, cognitive processing and motor behaviour of all living organisms, from amoebae to humans[1–4]. Neural circuit mechanisms based on neuronal and synaptic properties have been shown to process temporal information over the range of tens of microseconds to hundreds of milliseconds[5–7]. How neural circuits process temporal information in the range of seconds to minutes is much less understood. Studies of working memory in monkeys and rats have shown that neurons in the prefrontal cortex[8–10], the parietal cortex[9,11] and the thalamus[12] exhibit ramping activities that linearly correlate with the lapse of time until the end of a specific time interval of several seconds that the animal is trained to memorize. Many organisms can also memorize the time interval of rhythmic sensory stimuli in the timescale of seconds and can coordinate motor behaviour accordingly, for example, by keeping the rhythm after exposure to the beat of music. Here we report a form of rhythmic activity among specific neuronal ensembles in the zebrafish optic tectum, which retains the memory of the time interval (in the order of seconds) of repetitive sensory stimuli for a duration of up to ~20 s. After repetitive visual conditioning stimulation (CS) of zebrafish larvae, we observed rhythmic post-CS activities among specific tectal neuronal ensembles, with a regular interval that closely matched the CS. Visuomotor behaviour of the zebrafish larvae also showed regular post-CS repetitions at the entrained time interval that correlated with rhythmic neuronal ensemble activities in the tectum. Thus, rhythmic activities among specific neuronal ensembles may act as an adjustable 'metronome' for time intervals in the order of seconds, and serve as a mechanism for the short-term perceptual memory of rhythmic sensory experience.

The zebrafish tectum processes visual information and integrates it with inputs from other sensory modalities[13,14]. To investigate the ensemble neuronal activity triggered by visual stimulation, we used two-photon fluorescence imaging[15] of Ca^{2+} dynamics[16] to monitor the neuronal activities of a large population of cells (~200) simultaneously in intact, unanesthetized and unparalysed zebrafish larvae (5–14 days post-fertilization (d.p.f.), Fig. 1a, ref. 17). The amplitude of Ca^{2+} transients increases in individual neurons correlated with the number of spikes (Supplementary Fig. 1; see refs 17–20). Repetitive visual stimulation of the contralateral eye with a moving light bar across the visual field induced reliable responses in some tectal cells, but caused sporadic or habituating responses in others (Fig. 1b). Moreover, moving the bar stimuli in opposite directions activated different, but partially overlapping, neuronal ensembles (Fig. 1c and Supplementary Fig. 2a). The mean amplitudes of Ca^{2+} transients evoked in each neuron by consecutively moving bars in the same direction were highly correlated, whereas those induced by bars moving in opposite directions showed much lower correlation, suggesting

that the tectal-ensemble-evoked responses are stimulus-pattern-specific (Supplementary Fig. 2b, c).

In the absence of visual stimulation, synchronized Ca^{2+} transients among different tectal cells were rarely observed and showed no apparent regularity. However, after repetitive stimulation of the contralateral eye with a moving bar stimulus 20 times, and with an interstimulus interval (ISI) of 6 s (CS), we observed post-CS synchronized Ca^{2+} transients in a subpopulation of tectal neurons, at a time corresponding to the multiples of the ISI of the CS (Fig. 2a). This was also shown by the average profile (Fig. 2b) and onset time (Fig. 2c, d) of the Ca^{2+} transients for all cells. In general, 'rhythmic' Ca^{2+} transients occurred for up to three cycles (18 s) and were found in a subpopulation of neurons that were responsive to the CS, spatially dispersed
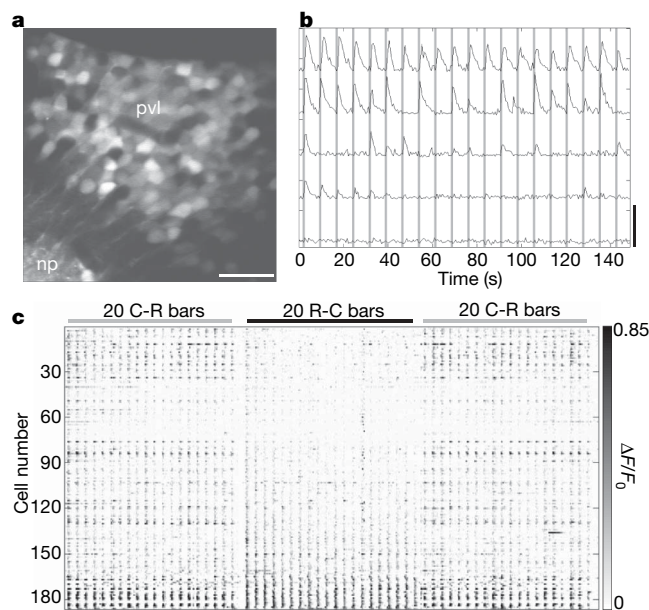

Figure 1 | Visual-stimulation-evoked Ca^{2+} transients in an ensemble of tectal neurons. a, An optical section of zebrafish larva tectum (6 d.p.f.), labelled with Ca^{2+} indicator and imaged using a two-photon microscope. np, neuropil; pvl, periventricular layer; scale bar, 20 μm. **b,** Five example neurons showing evoked Ca^{2+} transients in response to 20 repetitions of the moving bar (vertical grey bars) at an ISI of 6 s. Scale bar, 0.5 (change in baseline fluorescence, $\Delta F/F_0$). **c,** Raster plot of Ca^{2+} transients observed in 186 tectal neurons evoked by the moving bar at an ISI of 6 s. Grey, caudo-rostral (C-R); black, rostro-caudal (R-C). Neurons are sorted and numbered by their average evoked responses to the rostro-caudal stimulation. Grey scale represents the amplitude of Ca^{2+} transients in $\Delta F/F_0$. Scale bar, 20 s.

[1]Division of Neurobiology, Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. [2]Department of Physiology, University of California, San Francisco, California 94158, USA. †Present address: Laboratoire de Neurobiologie, UMR 8544, École Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France.
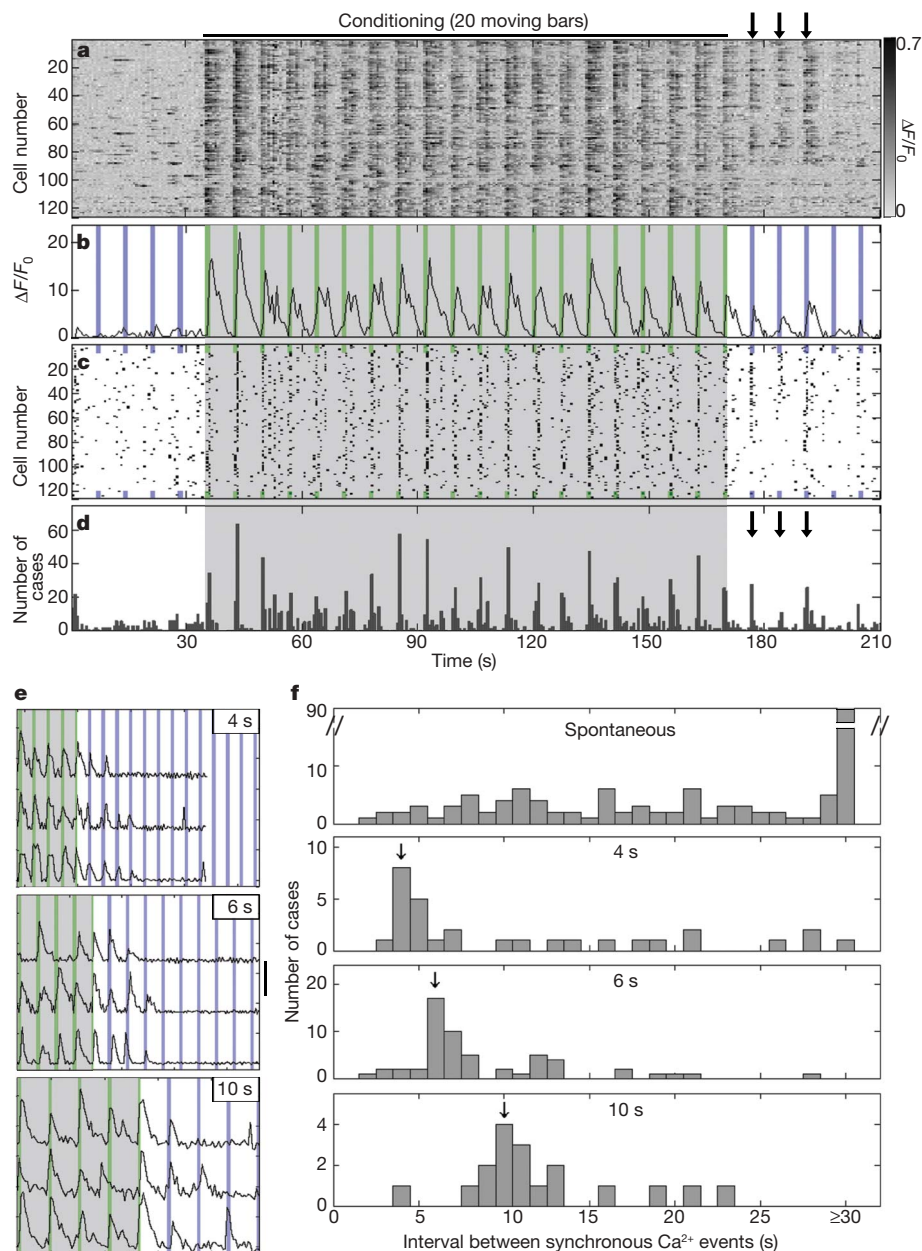
Figure 2 | Repetitive conditioning stimulation induces post-CS rhythmic Ca²⁺ transients. a, Raster plot of Ca²⁺ transients for 126 tectal neurons before, during and after CS (20 moving bar stimuli, 6-s ISI). Note that after the CS, synchronous activity was observed at a regular interval corresponding to the ISI of the CS (arrows) for a subset of neurons (arranged as 1 to 89). **b**, The average of all Ca²⁺ transients in **a**. The green bar denotes the moving bar stimulus; the blue bar represents the expected time corresponding to the ISI (6 s) of the CS should the CS start earlier or continue for further cycles. **c**, The onset time of Ca²⁺ transients for the data shown in **a**. **d**, Onset time histograms for data shown in **c**. The shaded area represents the CS period. **e**, Examples of Ca²⁺ transients evoked by the last five conditioning stimuli (ISIs of 4, 6 or 10 s; three neurons each), and observed during the post-CS period. Vertical scales ($\Delta F/F_0$) = 0.3. **f**, Histograms of inter-event intervals for synchronous Ca²⁺ events (see Methods for definition) in the absence of any sensory stimulation ('spontaneous', total time 130.5 min, from 35 trials), and for synchronous Ca²⁺ events during the first 30 s after CS at 4-, 6- or 10-s ISIs ($n = 16$, 38 or 17, respectively). Arrows represent the CS ISI.

across the ensemble (Supplementary Fig. 3a). Altogether, post-CS rhythmic activities were observed in 53 out of 110 experiments (48%, 23 larvae, 20 moving bars, 6 s ISI). The average amplitude of the integrated Ca²⁺ signals associated with rhythmic activities and their probability of occurrence were lower than that of the CS-evoked responses ($P < 0.001$, t-test, Fig. 3b), and they decayed with time (Supplementary Fig. 3b). Such rhythmic activities were not observed in the 3-d.p.f. larvae ($n = 8$) and seemed to emerge in the 4-d.p.f. larvae (one out of seven). Furthermore, the minimal number of stimuli for inducing post-CS rhythmic activities was ~10, and the robustness of these activities (the mean number of rhythmic cycles) increased with

the number of stimuli until it reached a plateau at ~50–100 cycles (Supplementary Fig. 4a, b). Uninterrupted rhythmic stimulation was required for inducing post-CS rhythmic activities, as shown by the lack of an apparent cumulative effect for two to four CS episodes (each consisting of ten stimuli) that were presented with 3-min spacing (Supplementary Fig. 4c). Similar post-CS rhythmic activities were also induced by CS other than moving light bars, including moving dark bars, looming light circles and wide-field light flashes (Supplementary Fig. 5).

Further experiments showed that the moving-bar CS of different ISIs (4, 6 and 10 s) induced post-CS rhythmic activities at the
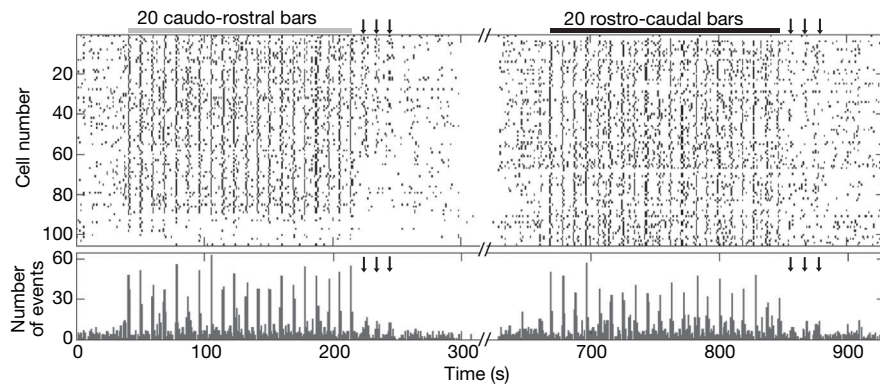
Figure 3 | Rhythmic activities of neuronal ensembles are stimulus-specific. The onset time of Ca^{2+} transients in an ensemble of 105 tectal neurons in response to a CS of 20 caudo-rostral (grey) and 20 rostro-caudal (black) moving bars, with neurons showing rhythmic activities entrained by the caudo-rostral CS (sorted as 1–65) and those un-responsive to the caudo-rostral CS (as 91–105). The two CS trials were separated by 5 min (marked by '//'). The bottom histograms represent the onset times of the Ca^{2+} transients shown at the top. Arrows depict the post-CS rhythmic activities.

conditioning ISI (Fig. 2e). The temporal precision of these entrained rhythmic activities were analysed for all experiments by plotting the histogram of the onset time of Ca^{2+} transients relative to that of the visual stimulus during the CS, or relative to the expected onset time should the CS started earlier or continued for further cycles. The distribution of the onset times was largely uniform before CS, but became clustered during the CS. Post-CS Ca^{2+} transients showed non-uniform onset time distributions for up to three cycles, with clear peaks in the first cycle for all three ISIs, and peaks up to the second and third cycle for the ISIs of 6 and 4 s, respectively (Supplementary Fig. 6). Because rhythmic activities were observed for up to $\sim$20 s for all ISIs, these activities may be limited by the time lapsed rather than by the number of entrained rhythmic cycles. The analysis of spontaneous ensemble Ca^{2+} transients showed that synchronous Ca^{2+} transients among an ensemble of tectal neurons occurred with a very low frequency ($\sim$0.02 Hz), and were rather uniformly distributed (Fig. 2f, 'spontaneous'). However, during the first 30 s after CS, the temporal distribution of synchronous Ca^{2+} events showed clear peaks at the entrained interval (Fig. 2f, '4, 6 or 10 s').

The post-CS rhythmic activities in the tectum are specific to the CS. Consecutive presentations (5 min apart) of two different types of CS—a light bar moving either caudo-rostrally or rostro-caudally, for 20 repetitions (ISI 6 s)—induced post-CS rhythmic Ca^{2+} transients in two partially overlapping tectal cell populations, which were essentially all within the ensemble that responded to the respective CS. In the experiment shown in Fig. 3, 54 out of the 67 neurons that responded to both CSs had post-CS rhythmic activities, and among these neurons, many (36 out of 54) showed rhythmic activities induced only by one CS (Supplementary Fig. 7a). Data from four other experiments showed similar results. Furthermore, rhythmic activities occurred in neurons that were not necessarily among those that were highly responsive to the stimulus, because there was no clear correlation between the mean amplitude of stimulus-evoked and the post-CS rhythmic Ca^{2+} transients of these neurons ($R^2 = 0.09$, Supplementary Fig. 7b). Thus, rhythmic Ca^{2+} transients probably reflect the activity of a specific neuronal ensemble entrained by the CS, rather than the activation of neurons with higher excitability. Post-CS rhythmic activity is unlikely to originate from the retina, because it was absent in the axon terminals of retinal ganglion cells expressing the genetically encoded Ca^{2+} indicator G-CaMP[21] (32 trials in 5 larvae, Supplementary Fig. 8 and Supplementary Methods).

To investigate the physiological relevance of CS-induced rhythmic activities of tectal neurons, we examined the visuomotor behaviour of zebrafish larvae (7–14 d.p.f.). The heads of the larvae were immobilized in agarose and tail kinematics was analysed before, during and after the visual CS. Repetitive wide-field light flashes evoked 'tail-flip'

behaviour (Fig. 4a, b) with a probability of $\sim$0.6. Notably, after 20 flashes of CS at an ISI of 4, 6 or 10 s, 30% of the larvae (24 out of 81, 13 larvae) showed post-CS tail flips in the absence of sensory stimulation for at least one cycle at the entrained ISI for a period of up to $\sim$20 s (Fig. 4c, d and Supplementary Movie 1). Spontaneous tail flips occurred at a low frequency ($\sim$0.015 Hz), without showing a preference for any specific interval (Fig. 4e, 'spontaneous'). Nevertheless, during the first 30 s after the CS, the distribution of tail-flip events clustered around the entrained interval (Fig. 4e, '4, 6 or 10 s').

By monitoring tectal ensemble Ca^{2+} transients and tail-flip behaviour simultaneously (Supplementary Fig. 9), we found a high correlation between the tail flip and the synchronous Ca^{2+} event in the tectal ensemble during the CS and the first 30 s after the CS, but a significantly lower correlation during the 60 s before, and the 31–60 s after, the CS (Fig. 4f and Supplementary Fig. 5). Moreover, normalized mean Ca^{2+} transients of the entire tectal ensemble during the first 30-s post-CS period were significantly higher for those associated with the rhythmic tail flip than those not associated with it (Fig. 4g and Supplementary Fig. 5).

The close resemblance and the correlation between entrained rhythmic tectal activities and post-CS tail flips suggest that entrained rhythmic activities contribute to short-term perceptual memory of visual experience. These rhythmic activities may raise the activity of specific neuronal ensembles at the entrained interval to a level closer to, and occasionally surpassing, that required for triggering the rhythmic activities, perhaps through the tectal outputs to the hindbrain[22,23]. This idea was further supported by the observation that the subthreshold stimuli that were normally ineffective in eliciting a tail flip (probability 0.06) had a significantly facilitating effect on evoking tail flip after the CS (20 flashes, ISI of 6 s) when they were presented at the entrained interval time. For the first four cycles after the CS, the probability of rhythmic tail flips was increased from 0.30 (24 out of 81, 13 larvae) to 0.49 (18 out of 37 trials, 7 larvae, $P = 0.045$, chi-squared test), but only when the subthreshold stimuli were applied in phase with the entrained interval. In contrast, the same subthreshold stimuli applied in antiphase (at 3, 9, 15 and 21 s post-CS) rarely evoked a tail flip (1 out of 14 trials, 4 larvae), and the probability of rhythmic tail flips at the entrained interval (0.29, 4 out of 14 trials, 4 larvae) was similar to that which was observed with no subthreshold stimulation ($P = 0.99$, chi-squared test). Furthermore, during the late phase of the CS, tail-flip behaviour was sometimes initiated shortly before the light stimulus onset in an 'anticipatory manner' (Fig. 4d and Supplementary Movie 1). Thus, the rhythmic activity may increase the sensitivity of the neural circuit to a specific sensory stimulus occurring at the entrained time interval.
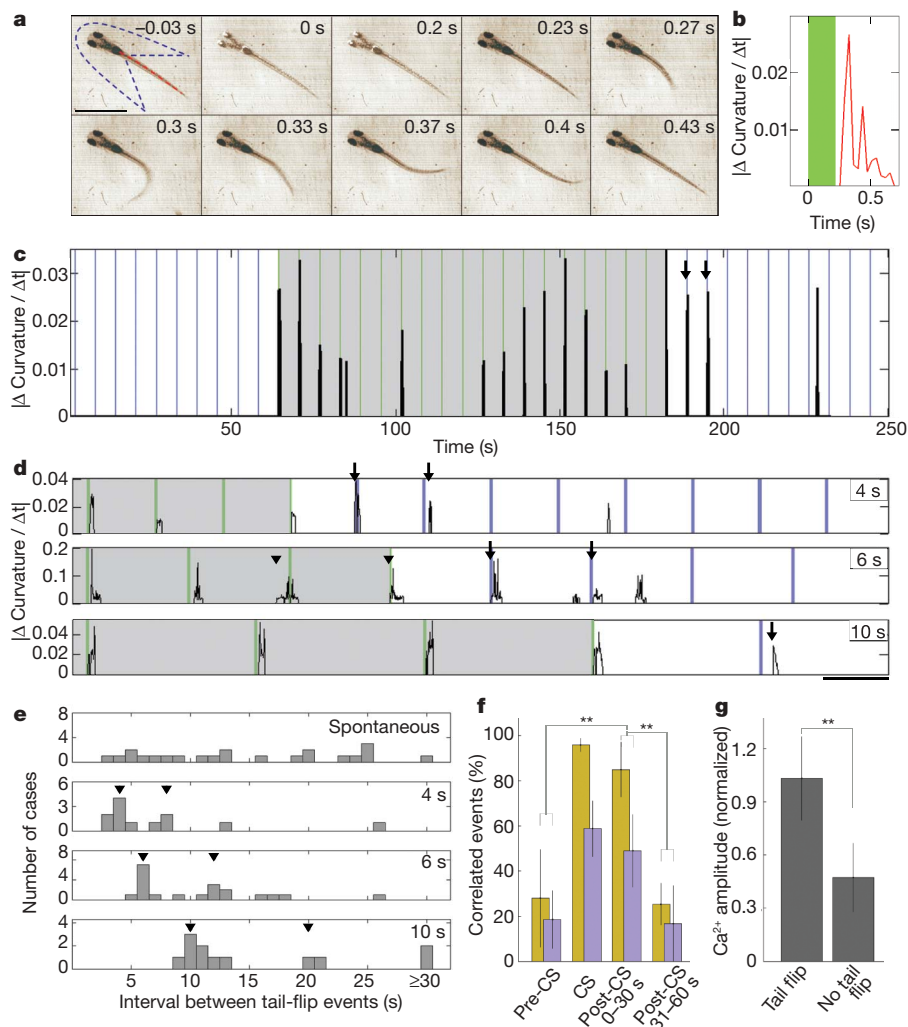
Figure 4 | Repetitive visual CS induces post-CS rhythmic motor behaviour.
a, Images depict the tail-flip behaviour of a head-immobilized zebrafish larva, evoked by a light flash (duration 0.2 s). Red dashed line represents the tail backbone curve; blue dashed line represents the contour of the restraining agarose. Scale bar, 1 mm. **b**, The absolute value of the average time (t) derivative of the tail backbone curvature, calculated from the images in **a**. Green bar represents the stimulus duration. **c**, Data from one example experiment before, during and after visual stimulation (20 flashes, 6-s ISI). Shaded area shows the CS period; arrows represent onset of the post-CS tail flips. Green lines show the CS ISI; blue lines denote the entrained interval time. **d**, Tail flips triggered by the last four CS stimuli of 4-, 6- or 10-s ISIs, and post-CS tail flips in the absence of sensory stimulation. Scale, 4 s. Arrowheads show tail flips initiated shortly before the stimulus. **e**, Histograms of inter-tail-flip event intervals in the absence of sensory

stimulation ('spontaneous', total time 30 min, from 15 trials), and for tail-flip events during the first 30 s after CS at 4-, 6- or 10-s ISIs ($n = 7$, 8 and 5, respectively). Arrowheads represent the CS ISI and multiples. **f**, The percentage of tail flips that correlated with synchronous Ca^{2+} events (yellow) and the percentage of synchronous Ca^{2+} events that correlated with tail flips (magenta) during the period corresponding to 60-s pre-CS, CS, and 0–30-s (or 31–60-s) post-CS ($n = 10$, 8 larvae). Percentages for the post-CS 0–30-s group were significantly different from those of the corresponding pre-CS and post-CS 31–60-s groups ($P < 10^{-3}$, non-parametric analysis of variance (ANOVA); **$P < 0.01$, Kolmogorov–Smirnov tests), but not from those of the CS group. **g**, The mean Ca^{2+} transients during the first 30-s post-CS period (normalized by the mean CS value), for all cases associated or not associated with post-CS rhythmic tail flips (**$P = 0.02$, Kolmogorov–Smirnov test, $n = 10$ trials, 8 larvae). Error bars, s.d.

Previous studies[24–28] have shown that repetitive visual stimulation at a regular ISI can result in spontaneous neuronal activity roughly at the expected time interval during and after the end of stimulation—a phenomenon known as 'omitted stimulus potential'. The omitted stimulus potential originates from the retina[24,25,28] and occurs for repetitive stimuli with much shorter ISIs (15–500 ms). It was found in cells not necessarily responsive to the stimuli[28] and has not been observed beyond one cycle[24–28]. The present results extend the phenomenon of the omitted stimulus potential by showing that the rhythmic activity of tectal circuits may act as an adjustable circuit 'metronome' that can be set to memorize stimulus time intervals in the order of seconds for a duration up to ~20 s, enabling the zebrafish larvae to estimate the time of a specific impending stimulus. Thus, short-term memory of rhythmic sensory experience may be

represented by entrained rhythmic activities, the neural circuit basis of which remain to be determined.

METHODS SUMMARY

1. Gibbon, J., Malapani, C., Dale, C. L. & Gallistel, C. R. Toward a neurobiology of temporal cognition: Advances and challenges. *Curr. Opin. Neurobiol.* **7,** 170–184 (1997).
2. Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of interval timing. *Nature Rev. Neurosci.* **6,** 755–765 (2005).
3. Buonomano, D. V. The biology of time across different scales. *Nature Chem. Biol.* **3,** 594–597 (2007).
4. Saigusa, T., Tero, A., Nakagaki, T. & Kuramoto, Y. Amoebae anticipate periodic events. *Phys. Rev. Lett.* **100,** 018101 (2008).
5. Carr, C. E. Processing of temporal information in the brain. *Annu. Rev. Neurosci.* **16,** 223–243 (1993).
6. Buonomano, D. V. Timing of neural responses in cortical organotypic slices. *Proc. Natl Acad. Sci. USA* **100,** 4897–4902 (2003).
7. Buonomano, D. V. & Karmarkar, U. R. How do we tell time? *Neuroscientist* **8,** 42–51 (2002).
8. Rainer, G., Rao, S. C. & Miller, E. K. Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* **19,** 5493–5505 (1999).
9. Quintana, J. & Fuster, J. M. From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex* **9,** 213–221 (1999).
10. Brody, C. D., Hernandez, A., Zainos, A. & Romo, R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* **13,** 1196–1207 (2003).
11. Leon, M. I. & Shadlen, M. N. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38,** 317–327 (2003).
12. Komura, Y. *et al.* Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature* **412,** 546–549 (2001).
13. Meek, J. Functional anatomy of the tectum mesencephali of the goldfish. An explorative analysis of the functional implications of the laminar structural organization of the tectum. *Brain Res.* **287,** 247–297 (1983).
14. Vanegas, H. & Ito, H. Morphological aspects of the teleostean visual system: a review. *Brain Res.* **287,** 117–137 (1983).
15. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248,** 73–76 (1990).
16. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100,** 7319–7324 (2003).
17. Niell, C. M. & Smith, S. J. Functional imaging reveals rapid development of visual response properties in the zebrafish tectum. *Neuron* **45,** 941–951 (2005).
18. McLean, D. L., Fan, J., Higashijima, S., Hale, M. E. & Fetcho, J. R. A topographic map of recruitment in spinal cord. *Nature* **446,** 71–75 (2007).
19. Ramdya, P., Reiter, B. & Engert, F. Reverse correlation of rapid calcium signals in the zebrafish optic tectum *in vivo*. *J. Neurosci. Methods* **157,** 230–237 (2006).
20. Yaksi, E. & Friedrich, R. W. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. *Nature Methods* **3,** 377–383 (2006).
21. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19,** 137–141 (2001).
22. Gahtan, E., Tanger, P. & Baier, H. Visual prey capture in larval zebrafish is controlled by identified reticulospinal neurons downstream of the tectum. *J. Neurosci.* **25,** 9294–9303 (2005).
23. Sato, T., Hamaoka, T., Aizawa, H., Hosoya, T. & Okamoto, H. Genetic single-cell mosaic analysis implicates ephrinB2 reverse signaling in projections from the posterior tectum to the hindbrain in zebrafish. *J. Neurosci.* **27,** 5271–5279 (2007).
24. Bullock, T. H., Hofmann, M. H., New, J. G. & Nahm, F. K. Dynamic properties of visual evoked potentials in the tectum of cartilaginous and bony fishes, with neuroethological implications. *J. Exp. Zool.* **256,** 142–155 (1990).
25. Bullock, T. H., Hofmann, M. H., Nahm, F. K., New, J. G. & Prechtl, J. C. Event-related potentials in the retina and optic tectum of fish. *J. Neurophysiol.* **64,** 903–914 (1990).
26. Bullock, T. H., Karamursel, S. & Hofmann, M. H. Interval-specific event related potentials to omitted stimuli in the electrosensory pathway in elasmobranchs: an elementary form of expectation. *J. Comp. Physiol. [A]* **172,** 501–510 (1993).
27. Bullock, T. H., Karamursel, S., Achimowicz, J. Z., McClune, M. C. & Basar-Eroglu, C. Dynamic properties of human visual evoked and omitted stimulus potentials. *Electroencephalogr. Clin. Neurophysiol.* **91,** 42–53 (1994).
28. Schwartz, G., Harris, R., Shrom, D. & Berry, M. J. II. Detection and prediction of periodic patterns by the retina. *Nature Neurosci.* **10,** 552–554 (2007).
29. Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L. & Raible, D. W. *nacre* encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development* **126,** 3757–3767 (1999).
30. O'Malley, D. M. *et al.* Optical physiology and locomotor behaviors of wild-type and nacre zebrafish. *Methods Cell Biol.* **76,** 261–284 (2004).

METHODS

Zebrafish preparation. Embryos from wild-type zebrafish and zebrafish nacre, with the latter lacking melanophores[29,30], were collected and raised at 28.5 °C in E3 embryo medium[31]. The larvae were kept under 14/10 h on/off light cycles and were fed after 6 d.p.f. All experiments were approved by the University of California Berkeley's Animal Care and Use Committee.

Calcium imaging of ensemble neuronal activities. Calcium indicator dye Oregon Green 488 BAPTA-1 AM was dissolved in dimethylsulphoxide with 20% pluronic (10 mM) and further diluted 10:1 in Evan's solution[32]. For the dye loading into tectal neurons, larvae at 5–14 d.p.f. were embedded in 1.2% low-melting-point agarose and anaesthetized with 0.02% MS-222 or Evan's solution at ~13 °C, similar to methods described previously[17,32]. The larvae were bolus injected under a stereomicroscope, using pressure injection (2–3 pulses of 30 ms duration, at ~10 pounds per square inch) through a micropipette (tip opening 2–3 μm). Larvae were incubated in the dark (at ~24 °C) in E3 for 1 h before use. Ca^{2+} imaging was performed mainly in the periventricular layer of the optic tectum, using either a Zeiss confocal (at 488 nm) or a custom-built two-photon microscope system (at 790 nm), with ×40 water-immersion objective (NA 0.8). Continuous scanning (1–2.7 Hz) was triggered by the visual stimulation software. Owing to pigmentation, wild-type larvae were imaged using the confocal system. Under low laser power (<2 mW), some tectal cells weakly responded to the onset of the 488 nm laser light. We thus discarded data collected during the first 5-s period after the laser onset.

Visual stimulation. Custom-made software (written in Matlab and psychtoolbox[33,34]) were used to drive light bars to move in various directions (constant speed $65° s^{-1}$, duration 1 s), and whole-field looming or flashing lights of various durations. Standard CS consisted of 20 repetitions with regular ISIs, defined as the time between the offset of the previous and the onset of the next stimulus. Visual stimuli were applied using a 14 × 9 mm LCD screen, with green light filtered (by Kodak-32) to avoid interference with the Ca^{2+} dye signal. The larvae were mounted dorsal-side up on the edge of a platform in an E3 solution-filled chamber, allowing an unobstructed view of the screen. The screen (covering 90° × 65° of the visual field) was centred around the eye of the larvae, contralateral to the dye-loaded tectum and positioned 7 mm from the eye—a distance that allows both the positioning of the objective above the tectum, and the proper focusing of the stimulus on the retina[35].

Ca^{2+} imaging analysis. A series of images obtained for each trial were aligned to compensate for drifts in the *xy* plane, by minimizing the mean square difference of intensities between the first frame and the rest of the data set, using spline processing (TurboReg[36]). Data showing drifts in the *z* plane were discarded. Regions of interest (ROIs) corresponding to each of the imaged tectal neurons were manually marked on the average image calculated from the entire series, and the averaged pixel intensity within ROIs was calculated. The change in intensity of each ROI was calculated as $\Delta I = \frac{I - I(t)_{base}}{I(t)_{base}}$, in which I is the average intensity of the ROI and I_{base} is a polynomial function of second or third degree fitted to the whole trace, demarking the trace's baseline. The polynomial function also served to correct for slight baseline changes due to photobleaching. A Ca^{2+} transient was considered an activity event when it surpassed a level of 1.5 s.d. above the baseline average and had a typical profile of fast rise and slow decay (as determined by visual inspection). Cells showing typical glial morphology (with triangular soma and long thick projections) or Ca^{2+} transients of prolonged decay time (>5 s) were excluded from further analysis. A 'synchronous Ca^{2+} event' among neurons in the ensemble is considered to have occurred when the onset histograms of Ca^{2+} transients (Fig. 2d) showed a peak that surpassed the threshold of 1 s.d. above the average. A given cell (or ensemble) was considered as showing 'entrained rhythmic activities' when at least one Ca^{2+} transient (or synchronous Ca^{2+} event) fell within ±0.5 s around the multiples of the ISI of the CS.

Behavioural assay. Wild-type and nacre larvae (7–14 d.p.f.) were embedded in agarose and submerged in E3 medium in the recording chamber. The agarose around the tail was removed to allow escape/swimming behaviours with kinematics similar to those of free behaviour[37]. Only larvae showing low frequency of spontaneous motor behaviours and reliable visuomotor behaviour (>50% success rate) without habituation were used (Supplementary Movie 1 and ref. 38). Experiments were performed under low ambient light and filmed at 60 Hz. Whole-field light flashes (duration 200 ms) were used as the CS instead of moving-bar stimuli because they more reliably elicited visuomotor responses. To determine the onset and duration of the tail flip, we fitted a backbone curve along the midline of the larva's tail and calculated the absolute value of the average value of its curvature (the derivative of the tangent angle with respect to arc length) for all time frames. Subthreshold stimulation was defined as the maximal light intensity that consistently failed to evoke motor responses. This intensity varied between larvae and was determined 5 min before the experiment.

Simultaneous recording of Ca^{2+} dynamics and motor behaviour. Zebrafish larvae (7–15 d.p.f.) were restrained in 2% agarose except for the last third of the tail. The larva was then placed under an upright confocal microscope and illuminated with infrared LEDs. Larva behaviour was monitored with a ×20 custom-made mini-microscope (connected to a video camera and shielded with a 488 nm notch filter; Supplementary Fig. 7). Synchronized acquisition of both fluorescence and bright-field images was done by custom software (Matlab). Data showing movement artefacts were discarded. Synchronous Ca^{2+} events (see definition described previously) and tail-flip behaviours were considered correlated if the synchronous Ca^{2+} event fell within a 1-s time window (±0.5 s around the onset time of the tail flip).

31. Westerfield, M. *The Zebrafish Book; A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (Eugene, Univ. Oregon Press, 2000).
32. Brustein, E., Marandi, N., Kovalchuk, Y., Drapeau, P. & Konnerth, A. ''In vivo'' monitoring of neuronal network activity in zebrafish by two-photon Ca^{2+} imaging. *Pflugers Arch.* **446**, 766–773 (2003).
33. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
34. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
35. Easter, S. S. Jr & Nicola, G. N. The development of vision in the zebrafish (*Danio rerio*). *Dev. Biol.* **180**, 646–663 (1996).
36. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
37. Ritter, D. A., Bhatt, D. H. & Fetcho, J. R. *In vivo* imaging of zebrafish reveals differences in the spinal networks for escape and swimming movements. *J. Neurosci.* **21**, 8956–8965 (2001).
38. Burgess, H. A. & Granato, M. Modulation of locomotor activity in larval zebrafish during light adaptation. *J. Exp. Biol.* **210**, 2526–2539 (2007).

LETTERS

Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases

Pilar Redondo[1]*, Jesús Prieto[2]*, Inés G. Muñoz[1], Andreu Alibés[3], Francois Stricher[3], Luis Serrano[3,4], Jean-Pierre Cabaniols[5], Fayza Daboussi[5], Sylvain Arnould[5], Christophe Perez[5], Philippe Duchateau[5], Frédéric Pâques[5], Francisco J. Blanco[2]† & Guillermo Montoya[1]

Xeroderma pigmentosum is a monogenic disease characterized by hypersensitivity to ultraviolet light. The cells of xeroderma pigmentosum patients are defective in nucleotide excision repair, limiting their capacity to eliminate ultraviolet-induced DNA damage, and resulting in a strong predisposition to develop skin cancers[1]. The use of rare cutting DNA endonucleases—such as homing endonucleases, also known as meganucleases—constitutes one possible strategy for repairing DNA lesions. Homing endonucleases have emerged as highly specific molecular scalpels that recognize and cleave DNA sites, promoting efficient homologous gene targeting through double-strand-break-induced homologous recombination. Here we describe two engineered heterodimeric derivatives of the homing endonuclease I-CreI, produced by a semi-rational approach. These two molecules—Amel3–Amel4 and Ini3–Ini4—cleave DNA from the human *XPC* gene (xeroderma pigmentosum group C), *in vitro* and *in vivo*. Crystal structures of the I-CreI variants complexed with intact and cleaved *XPC* target DNA suggest that the mechanism of DNA recognition and cleavage by the engineered homing endonucleases is similar to that of the wild-type I-CreI. Furthermore, these derivatives induced high levels of specific gene targeting in mammalian cells while displaying no obvious genotoxicity. Thus, homing endonucleases can be designed to recognize and cleave the DNA sequences of specific genes, opening up new possibilities for genome engineering and gene therapy in xeroderma pigmentosum patients whose illness can be treated *ex vivo*.

Meganucleases recognize large DNA sequences and cleave their cognate site without affecting genome integrity. However, new nucleases with customized specificity cannot be designed on the basis of amino acid sequence alone, because specific recognition is based on protein–DNA interactions, which can be identified only in the three-dimensional structure of the complex. A full understanding of the molecular basis of DNA recognition is therefore essential. Zinc-finger DNA-binding domains[2] were recently fused to the catalytic domain of the FokI endonuclease to induce recombination in various cell types, including human embryonic stem and lymphoid cells[3–6]. However, some of these chimaeras can be highly toxic to cells[3,7], probably due to their low level of specificity, which could be improved by rational design[8,9]. Given their high selectivity, homing endonucleases are ideal scaffolds to engineer accurate enzymes for DNA cleavage and recombination[10]. I-CreI is a member of the LAGLIDADG homing endonuclease family[11] with only one such motif[12] and functions as a homodimer. Crystal structures of I-CreI (ref. 13) show that each monomer contains its own DNA-binding region and that the catalytic centre is formed at the dimer interface.

We have developed a method for generating new DNA specificities on the basis of the engineering of each DNA-binding region to promote the recognition and cleavage of different DNA sequences[14,15]. Heterodimeric I-CreI variants were generated by combining a semi-rational approach with high-throughput methods for creating thousands of homodimeric I-CreI derivatives with local changes in specificity. The new variants display mutations clustered in two different DNA-binding subdomains, and cleave sequences differing from the palindromic I-CreI target at positions ± 8, ± 9 and ± 10 (10NNN) or ± 3, ± 4 and ± 5 (5NNN; Fig. 1a). A combinatorial strategy was used to assemble distinct clusters of mutations recognizing the 10NNN and 5NNN regions into globally engineered proteins with predictable specificities[15]. These combinatorial homodimeric mutants cleave a combined target comprised of a patchwork of the targets cleaved by the parental molecules. The monomers of these homodimeric
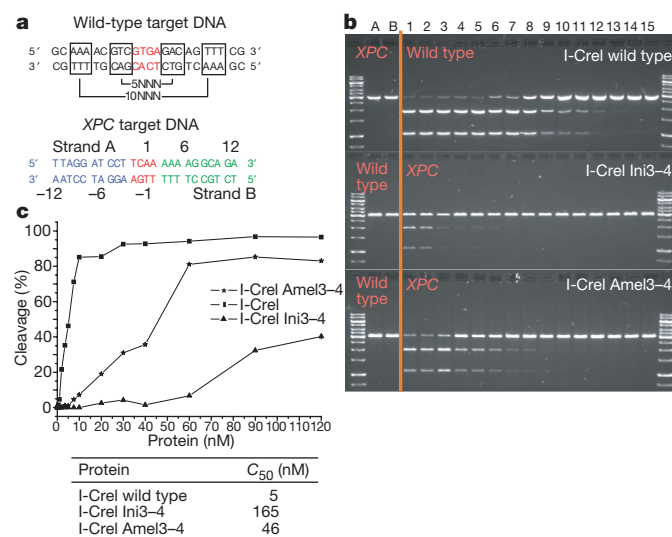

Figure 1 | New heterodimer cleavage activity and specificity. a, Wild-type and *XPC* target DNA sequences. The 5NNN and 10NNN regions are boxed. **b,** Target DNA cleavage *in vitro* by I-CreI and the heterodimers. The reaction mixture included a target concentration of 2 nM and a protein concentration of 120, 90, 60, 40, 30, 20, 10, 7.5, 5, 3.5, 2, 1, 0.5, 0.25 and 0 nM (lanes 1–15), or 120 nM (lane A) or 0 nM (lane B). Lanes A and B show wild-type DNA and lanes 1–15 show *XPC* DNA sequence. **c,** Representation of the *in vitro* cleavage assay and C_{50} (concentration of enzyme required to cleave 50% of the 2 nM target DNA).

[1]Macromolecular Crystallography Group, [2]NMR group, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), c/Melchor Fdez. Almagro 3, 28029 Madrid, Spain. [3]European Molecular Biology Laboratory (EMBL)-CRG Systems Biology Unit, Centre de Regulació Genòmica (CRG), Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain. [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), [5]CELLECTIS S.A., 102 Avenue Gaston Roussel, 93235 Romainville, France. †Present address: CIC bioGUNE, Parque Tecnológico de Vizcaya, Edificio 800, 48160 Derio, Spain.
*These authors contributed equally to this work.

mutants can be combined by coexpression to create heterodimeric proteins, thereby increasing the number of potential targets. The resulting library of mutants was used to design meganucleases for targeting a sequence of the *XPC* gene contained in the intron between exons 3 and 4 of the human *XPC* locus[16].

Screening in yeast yielded heterodimeric variants of I-CreI that cleaved DNA derived from the *XPC* gene but not with the high efficiency required for gene targeting in mammalian cells. This level of activity may be due to the *XPC* DNA sequence, which differs from the wild-type sequence not only at the 10NNN and 5NNN positions but also in the central four base pairs. The effect of these DNA positions on I-CreI activity is unclear. We therefore optimized these variants by random mutagenesis followed by screening for high levels of activity. The optimized variants have a mutated residue close to the active site and a few other mutations (Supplementary Table 1) not directly involved in DNA binding and catalysis. These enzymes efficiently cleaved the *XPC* target sequence, which differs in 17 bases from the wild type, and induced targeted recombination[16].

We have analysed two of these variants: Ini3–Ini4 generated by the combinatorial approach, and Amel3–Amel4 further optimized by random mutagenesis. The cleavage properties of both the homodimeric (Amel3–Amel3, Amel4–Amel4, Ini3–Ini3 and Ini4–Ini4) and the heterodimeric (Ini3–Amel4, Amel3–Ini4, Amel3–Amel4 and Ini3–Ini4) forms were analysed *in vitro* (Fig. 1b, c and Supplementary Fig. 1). Heterodimers were obtained by unfolding and refolding of an equimolar mixture of the corresponding homodimers (Supplementary Fig. 1). This process mimics coexpression in cells, and should generate a mixture of the two homodimers plus the heterodimer. Each homodimer was found to cleave its own combined palindromic target, but only mixtures with heterodimers cleaved the *XPC* target, with the most efficient enzymes being those containing one or both optimized monomers.

We isolated the Ini3–Ini4 and Amel3–Amel4 heterodimers after coexpression of the monomers with different affinity tags (Methods). The identity of the purified heterodimers was confirmed by mass spectrometry (Supplementary Fig. 2), and pull-down experiments showed that the homodimeric species were present in only very small amounts, if at all, confirming the stability of the heterodimers (Supplementary Fig. 3). Both heterodimers cleaved their corresponding targets less efficiently than I-CreI (Fig. 1b, c), but the optimized heterodimer was significantly more active than the initial one (Fig. 1c).

The crystal structures of Amel3–Amel4 and Ini3–Ini4 in complex with *XPC* DNA were solved by molecular replacement (Fig. 2a, b and Supplementary Table 2). The structures in the presence of Ca^{2+} and Mg^{2+} provided snapshots of the bound and cleaved states of catalysis, making it possible to compare these engineered variants with the wild type. The structures of I-CreI in complex with its target DNA crystallized with Ca^{2+} and Mg^{2+} were very similar in their overall conformation (0.36 Å Cα root mean squared deviation). Superimposition of the Ini3–Ini4 and Amel3–Amel4 heterodimers on the wild-type structures also resulted in a similar overall conformation (Cα root mean squared deviation of 0.48 Å to 0.57 Å, and 0.36 Å between them). However, there are conformational changes in the DNA-binding regions of the protein concomitant to changes in the DNA structure in the same area (Fig. 3). The loop between Lys 28 and Lys 36, which contains mutations Y33S and K28E, is displaced and forces local changes on the loop between Ala 115 and Asp 120 in its neighbourhood (Fig. 3a). These coordinated changes allow the mutations introduced into Amel3–Amel4 and Ini3–Ini4 to recognize the new DNA bases. The Y33S mutation in Amel3 and Ini3 induces the loss of the hydrogen bond between Tyr 33 and adenine-10 on the wild-type B strand (see Fig. 1a), and the appearance of new interactions between Ser 33 and cytosine-11 on the *XPC* B strand. The mutations K28E in the Ini4 and Amel4 monomers and K28A in Amel3 induce a change in the hydrogen bond pattern with DNA in the mutants with respect to the wild type on
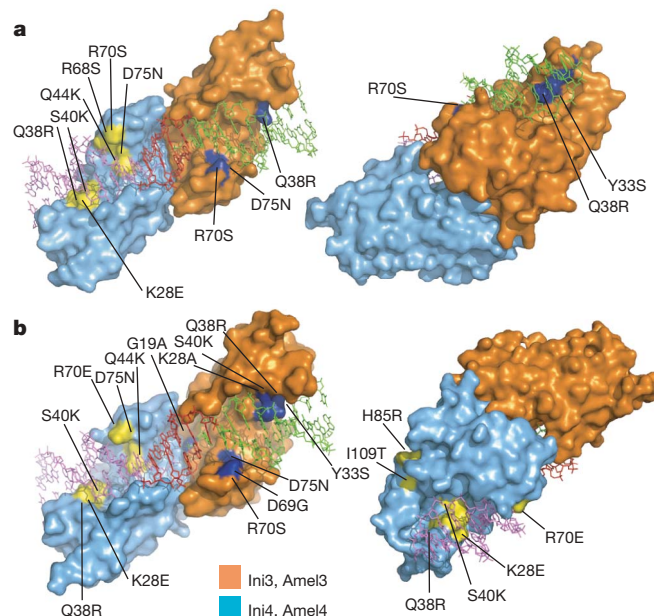
Figure 2 | Crystal structures of the Ini3–Ini4 and Amel3–Amel4 heterodimers. a, b, Surface models of the Ini3–Ini4 (**a**) and Amel3–Amel4 (**b**) structures in complex with *XPC* DNA. The DNA is coloured according to the binding regions of each monomer (violet for Amel4, Ini4; green for Amel3, Ini3; and red for the central four base pair). Mutations are mapped on the protein surface in yellow and dark blue for monomers Amel3, Ini3 and Amel4, Ini4, respectively (Supplementary Fig. 4).

both halves of the DNA target. Instead of forming a hydrogen bond with thymine −7 on the wild-type B strand, the Glu 28 interacts with cytosine −8 on the *XPC* B strand. These modifications occur in one
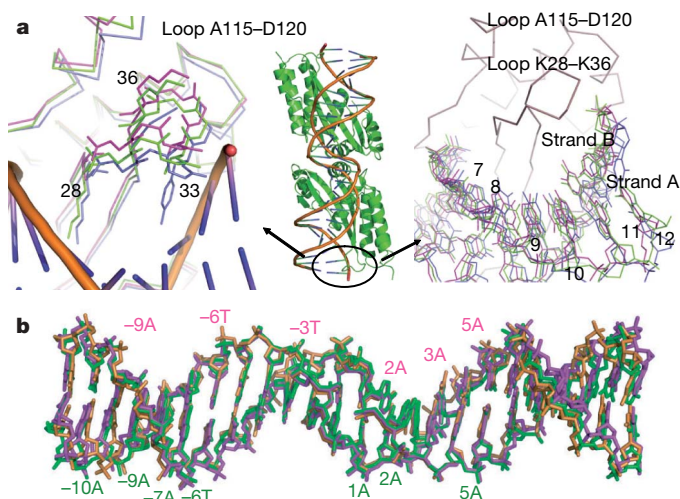
Figure 3 | Structural comparison of the wild-type, Amel3–Amel4 and Ini3–Ini4 complexes. a, Structural differences in the Lys 28–Lys 36 loop and 7–12 bases of the bound DNA between the Amel3 (magenta), Ini3 (green) and I-CreI (blue). The location of this region is indicated in ribbons in the central panel. The structural changes in the protein and DNA moieties are highlighted on both sides. Left, the Amel3, Ini3 and I-CreI proteins displaying amino acid side chains defining the loop and the mutated positions, including the Amel DNA molecule in a cartoon representation. Right, the DNA molecules of the Amel, Ini and I-CreI with the Cα trace of the Amel3 monomer (violet). Only one DNA and one protein are shown in both images. **b,** Superimposition of the DNA structures from the wild-type (orange), Amel3–Amel4 (magenta) and Ini3–Ini4 (green) complexes. Labelled bases make the largest contribution to the increase in energy when mutated to the wild-type sequence.

monomer whereas the wild-type interaction is conserved in the other one, introducing differences in binding of each monomer within the heterodimer. The changes in the Lys 28–Lys 36 loop also allow the Q38R mutation, present in all the monomers, to form a hydrogen bond through its guanidinium group with the purines guanine −9 on the *XPC* strand A and guanine −9 on the *XPC* B strand, which correspond to adenines in the wild type (Supplementary Information and Supplementary Figs 4–6).

The global analysis of the protein–DNA interactions in the heterodimers and its comparison to I-CreI indicate that the mutations on the different monomers generate an asymmetric DNA-binding environment on the I-CreI scaffold, facilitating the accommodation of a non-palindromic target such as *XPC* DNA. Interestingly, the initial selection yielded the Ini3–Ini4 that displayed a diminished hydrogen bond network with the DNA, which is recovered in the improved Amel3–Amel4 variant (Supplementary Table 3). Thus, the procedure seems to select initial variants that disrupt the DNA-binding symmetry of I-CreI, and the random mutagenesis seems to optimize the hydrogen bond network with the *XPC* target.

The residues at position 19, close to the active site, make van der Waals interactions with the phosphates of the DNA backbone and the base at position 2 in all of the structures. The G19A mutation promotes a less compact interaction at the carboxy termini of the $_{11}$LYLAGFVDGD$_{20}$ helices, which could improve the positioning of the catalytic residues[17] and/or favour the release of cleaved DNA.

The structures of the Amel and Ini active sites (Supplementary Fig. 5) suggest that their catalytic mechanism is similar to that of the wild-type enzyme, despite the mutations in the DNA-binding domains in both monomers, allowing cleavage of the phosphodiester bond between the bases in positions ±2 and ±3. No significant conformational change was observed in the A and B strands at the central four base pairs in the Amel3–Amel4 and Ini3–Ini4 structures. However, the main difference in this region between the *XPC* and wild type targets involves a transversion at position −2 on the wild-type strand A. This change in the DNA, in conjunction with mutations R70S and R70E, leads to changes in the position of the bases with respect to the protein in both heterodimers. This displacement fills the cavity generated by the purine–pyrimidine change.

To understand the energetics of these changes in the *XPC*-DNA-recognition mechanism, we performed an analysis of the protein–DNA contacts[18]. This study revealed that the Amel3–Amel4–*XPC* and Ini3–Ini4–*XPC* complexes could be exchanged without markedly affecting binding energy predictions. However, attempts to perform this *in silico* simulation with the wild-type target and wild-type structures failed (Supplementary Table 4), revealing the importance of the change in protein conformation for DNA recognition. These results suggest that the use of computer-assisted protein design to generate meganucleases with customized specificity for very different DNA sequences would require the support of different protein–DNA structures to manage the conformational diversity of protein and DNA.

An *in silico* analysis of the sequence–structure relationship of the DNA from I-CreI, Ini3–Ini4–*XPC* and Amel3–Amel4–*XPC* (Fig. 3b) structures in the absence of the protein moieties have been performed by modelling the wild-type sequence onto the *XPC* DNA structure and vice versa. The calculated differences in overall energy (Supplementary Table 5) showed that the wild-type sequence is energetically more compatible with the *XPC* DNA structure than the *XPC* sequence is with the wild-type DNA structure. These results suggest that some sequences may force the DNA to adopt a conformation energetically unfavourable for binding to a given I-CreI structure optimized for another DNA target.

The extensive redesign of the I-CreI protein resulted in a small loss of activity for Amel3–Amel4 and a much larger loss of activity for Ini3–Ini4 proteins (Fig. 1b, c). Because the aim of homing endonuclease design is to deliver tools for genomic engineering, it is essential to monitor the impact of this decrease on a real gene-targeting assay. We have previously described a chromosomal reporter system containing a meganuclease cleavage site for comparisons of the ability of different meganucleases to induce gene targeting in a similar chromosomal context[16] (Supplementary Information). Heterodimer targeting frequencies were compared with that for I-SceI, the standard for studies of double-strand break (DSB)-induced gene targeting, and I-CreI. CHO cell lines were constructed with a *lacZ* reporter system inserted into exactly the same chromosomal locus and differing only in meganuclease recognition site. The resulting cell lines, carrying the inactive *lacZ* gene, could be used for evaluating the efficiency of DSB-induced gene targeting—*lacZ* repair, with I-SceI, I-CreI or the Amel3–Amel4 and Ini3–Ini4 heterodimers cleaving the *XPC* target (Fig. 4a). The Amel3–Amel4 variant was nearly as efficient (2.2×10^{-3}) as the I-SceI (1.1×10^{-2}) and I-CreI (7.0×10^{-3}) enzymes used as standards[16]. In addition, the non-optimized Ini3–Ini4 variant induced substantial gene targeting, although less efficiently than the optimized Amel3–Amel4 heterodimer. In contrast, little or no signal was observed with the homodimers or in the absence of meganuclease. Finally, the Amel3-Ini4 and Amel4-Ini3 heterodimers, each containing one of the optimized monomers, had intermediate activity levels, consistent with their cleavage activity *in vitro* (Supplementary Fig. 1). Tagged versions of the Amel3–Amel4 variants were not expressed at higher levels than their Ini3–Ini4 counterparts in CHO cells, but were still more active in our gene-targeting assay (data not shown). This result shows that the improved activity of Amel3–Amel4 *in vivo* is not due to a higher expression and/or enhanced stability, but instead to its higher activity, consistent with the *in vitro* results.

To assess the toxicity levels of the improved heterodimer due to non-specific DSBs, we monitored the phosphorylation of H2AX histones (γ-H2AX) and their localization into nuclear foci at the DSB sites. MRC5 human cells, containing the endogenous *XPC* site, displayed an average number of γ-H2AX foci slightly above the background level (Fig. 4b, c) after transfection with Amel3–Amel4 under the same conditions used to induce recombination in the gene-targeting assay (Fig. 4a). In contrast, when we used either a first-generation zinc-finger-fused nuclease[9] or a I-CreI-derived meganuclease with low specificity[14,15], higher levels of γ-H2AX foci were detected (Fig. 4b, c). Therefore, the engineered Amel3–Amel4 meganuclease causes little, if any, off-site cleavage in human cells, demonstrating that it is a highly specific meganuclease comparable to I-SceI (refs 8, 9, 19). Similar results were obtained in the CHO-π10_XPC2 cell line (see Methods), in the parental cell line (CHO-K1; Supplementary Fig. 7) and in the XD17 cell line, a DSB-repair-deficient (*XRCC4⁻*) CHO-K1 cell line that has been shown to display more persistent foci after irradiation[20] (Supplementary Fig. 8).

Engineering a DNA-binding protein is a daunting challenge. Years of attempts to modify the substrate specificity of restriction enzymes[21,22] or recombinases[23–25] illustrate the difficulty of the task, and modular engineered zinc fingers have been the exception for a long time[26]. I-MsoI has recently been computationally redesigned to cleave a DNA sequence differing at two positions from its original target[27]. Our findings indicate that this approach might not be effective if extensive changes in the DNA sequence are required. However, our study also shows that, by combining rational engineering and high throughput screening, it is possible to entirely redesign the recognition properties of meganucleases, and potentially of other DNA-binding proteins, without altering activity and specificity. In patient's cells, the efficacy of DSB-induced gene targeting may depend on several factors including vectorization, cleavage activity, homologous recombination proficiency in the chosen cell type and, probably, the chromatin status of the targeted locus. In contrast, specificity will depend only on the intrinsic properties of the engineered meganuclease.

Stem cell research has shown that functional skin can be generated with as little as 1% epithelial stem cells[28]. The combination of this technology with our tailored endonucleases raises new possibilities for gene therapy in patients with xeroderma pigmentosum and other
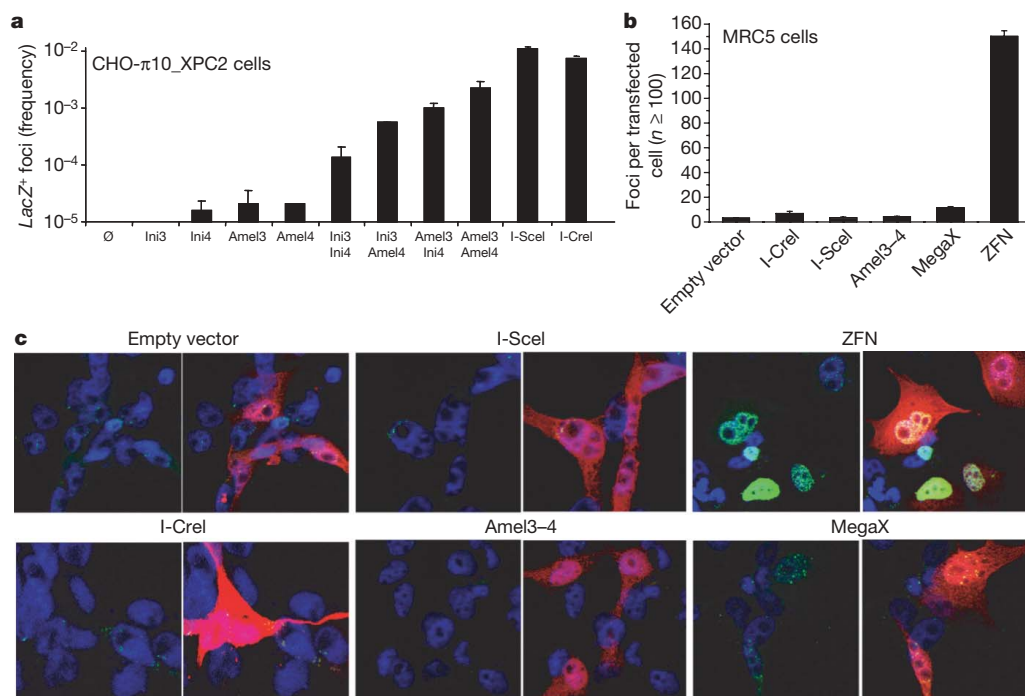
Figure 4 | Analysis of the efficiency of gene repair by customized meganucleases and genotoxicity controls. a, A cellular model was designed in CHO-K1 (CHO-π10_XPC2). *lacZ* gene function is restored when 2 μg of a *lacZ* repair matrix is cotransfected with 1 μg of a meganuclease expression vector(s) (see Methods). The cell line carrying the *XPC* recognition site was transfected with the *lacZ* repair matrix alone (Ø) or together with a meganuclease expression vector(s) (Ini3 and Ini4 homodimers, Amel3 and Amel4 homodimers, Ini3–Ini4, Ini3–Amel4, Amel3–Ini4, Amel3–Amel4 heterodimers, I-SceI and I-CreI). Error bars represent the s.d. of the measured values. **b**, γ-H2AX immunocytochemistry. Human MRC5 cells were transfected with 1 μg of plasmid(s) encoding meganuclease(s), or first-generation zinc-finger-fused nuclease (ZFN heterodimer, L^{wt}–R^{wt} as defined in ref. 9) targeting the human *IL2RG* locus[9] and 200 ng of plasmid encoding RFP (red fluorescent protein). Empty vector was used as control. The number of foci in more than 100 cells co-transfected with RFP was counted for each condition and the averaged number of three experiments is represented in the histogram including their s.e.m. MegaX is a I-CreI-derived meganuclease. **c**, View of a typical γ-H2AX foci in MRC5 cells transfected with meganucleases or zinc-finger-fused nuclease (L^{wt}−R^{wt}).

monogenic diseases, such as haematological diseases that could be treated *ex vivo*.

METHODS SUMMARY

Amel3, Amel4 and Ini3, Ini4 were cloned in the CDFDuet-1 plasmid for coexpression and the heterodimers were selectively purified using two different affinity tags. Crystallization screenings were performed using the hanging-drop method after heterodimer–*XPC* DNA complex formation. Diffraction data sets were collected using synchrotron radiation and the crystal structures were solved using the molecular replacement method.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cleaver, J. E. Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nature Rev. Cancer* **5**, 564–573 (2005).
2. Pabo, C. O., Peisach, E. & Grant, R. A. Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* **70**, 313–340 (2001).
3. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
4. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
5. Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003).
6. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature Biotechnol.* **25**, 1298–1306 (2007).
7. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169–1175 (2002).
8. Szczepek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnol.* **25**, 786–793 (2007).
9. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnol.* **25**, 778–785 (2007).
10. Jacquier, A. & Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383–394 (1985).
11. Chevalier, B. S. & Stoddard, B. L. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **29**, 3757–3774 (2001).
12. Wang, J., Kim, H. H., Yuan, X. & Herrin, D. L. Purification, biochemical characterization and protein–DNA interactions of the I-CreI endonuclease produced in *Escherichia coli*. *Nucleic Acids Res.* **25**, 3767–3776 (1997).
13. Jurica, M. S., Monnat, R. J. Jr & Stoddard, B. L. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell* **2**, 469–476 (1998).
14. Arnould, S. *et al.* Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.* **355**, 443–458 (2006).
15. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* **34**, e149 (2006).
16. Arnould, S. *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* **371**, 49–65 (2007).
17. Silva, G. H. & Belfort, M. Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res.* **32**, 3156–3168 (2004).
18. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
19. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nature Cell Biol.* **9**, 675–682 (2007).
20. Saintigny, Y., Delacote, F., Boucher, D., Averbeck, D. & Lopez, B. S. XRCC4 in G1 suppresses homologous recombination in S/G2, in G1 checkpoint-defective cells. *Oncogene* **26**, 2769–2780 (2007).
21. Rimseliene, R., Maneliene, Z., Lubys, A. & Janulaitis, A. Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *J. Mol. Biol.* **327**, 383–391 (2003).
22. Samuelson, J. C. & Xu, S. Y. Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity. *J. Mol. Biol.* **319**, 673–683 (2002).
23. Buchholz, F. & Stewart, A. F. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nature Biotechnol.* **19**, 1047–1052 (2001).

24. Santoro, S. W. & Schultz, P. G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl Acad. Sci. USA* **99,** 4185–4190 (2002).

25. Voziyanov, Y., Konieczka, J. H., Stewart, A. F. & Jayaram, M. Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. *J. Mol. Biol.* **326,** 65–76 (2003).

26. Jamieson, A. C., Miller, J. C. & Pabo, C. O. Drug discovery with engineered zinc-finger proteins. *Nature Rev. Drug Discov.* **2,** 361–368 (2003).

27. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441,** 656–659 (2006).

28. Bernerd, F. *et al.* Clues to epidermal cancer proneness revealed by reconstruction of DNA repair-deficient xeroderma pigmentosum skin *in vitro. Proc. Natl Acad. Sci. USA* **98,** 7817–7822 (2001).

METHODS

Construction of target clones. The wild-type (5′-TCAAAACGTCGTA-CGACGTTTTGA-3′) and XPC (5′-TTAGGATCCTTCAAAAAAGGCAGA-3′) DNA target sequences were cloned as follows: oligonucleotides containing the target site (Eurogentec) were amplified by PCR to generate double-stranded target DNA and then were cloned using the Gateway protocol (Invitrogen) into the mammalian reporter vector pcDNA3.1-LACZ, described previously[16] and containing an I-SceI target site as control. The same target sequences were also cloned into pGEM-T Easy vector (Promega). The XPC sequence used in this work corresponds to the A1 site described previously[3]. Resulting clones were verified by sequencing (MilleGen).

Cloning. The open reading frames (ORFs) of all the monomers (Amel3, Amel4, Ini3 or Ini4), described previously as M3, M4, I2 and I3, respectively[16], were cloned using the Gateway protocol (Invitrogen). ORFs were amplified by PCR of yeast DNA using the primers: B1F, 5′-GGGGACAAGTTTGTACAAAAAA-GCAGGCTTCGAAGGAGATAGAACCATGGCCAATACCAAATATAACAAA-GAGTTCC-3′, and B2R, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGT-TTAGTCGGCCGCCGGGGAGGATTTCTTCTTCTCGC-3′ from Eurogentec. PCR products were cloned in CHO gateway expression vector pCDNA6.2 from Invitrogen. Resulting clones were verified by sequencing (MilleGen).

Protein expression and purification. The mutant I-CreI proteins were cloned in the expression vector pET24-d(+) plasmid and the production of the corresponding homodimeric proteins was performed as described[29]. To coexpress the heterodimeric I-CreI derivatives (Amel3–Amel4 and Ini3–Ini4), one of the monomers' (either Amel3 or Ini4) ORF cleaved by NcoI and EcoRI was cloned into CDFDuet-1 vector (Novagen) with a 6× His tag at the C terminus. The other monomer's (either Amel4 or Ini3) ORF cleaved by NdeI and XhoI was cloned with a Strep-tag at the C terminus into the corresponding CDFDuet-1 vector. The expression and purification of the homodimers was performed as described[29]. The double-tagged heterodimers were overexpressed in *Escherichia coli* Rosetta(DE3)pLysS cells grown in Luria Bertani medium containing 50 µg ml^{-1} streptomycin at 37 °C for three hours after addition of 0.3 mM IPTG when the D_{600} was around 0.6–0.8. The bacterial pellet was resuspended and the cells disrupted by sonication in 50 mM sodium phosphate, pH 8.0, 300 mM NaCl and 5% glycerol including protease inhibitors (complete EDTA-free tablets, Roche). The lysate was cleared by centrifugation (20,000g for 1 h). The filtered supernatant was applied to a Co^{2+}-loaded HiTrap chelating HP column (GE Healthcare) and the protein was eluted using an imidazole gradient (0–0.5 M). Protein-rich fractions (determined by SDS–PAGE) were collected and loaded onto a 5 ml Strep-Tactin Superflow column (IBA) previously equilibrated with 100 mM Tris-HCl, pH 8.0, 150 mM NaCl. The sample was eluted in one step with the previous buffer plus 2.5 mM desthiobiotin. The purified protein was subsequently concentrated using an Amicon Ultra system equipped with 10 kDa cutoff filter and loaded onto a PD-10 desalting column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, pH 8.0, and 150 mM NaCl to remove the desthiobiotin. Afterwards, the protein was concentrated to 15 mg ml^{-1}, flash-frozen in liquid nitrogen and stored at −80 °C. Protein concentration was determined by absorbance at 280 nm. The purity of the samples was checked by SDS–PAGE and heterodimer formation by western blot using an anti-His or an anti-Strep-tag antibody. All the purified proteins were found to be folded with a similar structure as the wild type (by circular dichroism and NMR) and to be dimeric in solution (by analytical ultracentrifugation). Dynamic light scattering also indicated essentially monodisperse solutions (data not shown).

Heterodimer stability. To check the stability of the heterodimer, its dissociation and formation of homodimers was checked over a period of 3 days. 4.5 mg of Ini3–Ini4 and Amel3–Amel4 heterodimers (and one of their corresponding homodimers as controls) were incubated at 37 °C for 0, 1, 2, 4 and 72 h (0 h for the control homodimers). After incubation, the solutions were mixed with Talon resin for 25 min, centrifuged, and the supernatants loaded onto an 18% acrylamide–SDS gel to monitor the heterodimer dissociation. If homodimer formation takes place during that time the homodimer containing the Strep-tag should not bind to the resin and must stay in the supernatant appearing as a band in the gel. The absence of this band demonstrates the stability of the heterodimer over this period of time.

Heterodimer–DNA complex formation and crystallization. The 24-bp-long XPC target DNA was purchased from Proligo and consisted of two strands of sequence: 5′-TTAGGATCCTTCAAAAAAGGCAGA-3′ and 5′-TCTGCCTTT-TTTGAAGGATCCTAA-3′, which form a 24-bp blunt-end duplex on mixing, heating and cooling. The protein–DNA complex was obtained in the presence of either 2 mM CaCl$_2$ or MgCl$_2$ (to obtain the bound and cleaved states of the target DNA) by pre-warming the meganuclease and the oligonucleotide samples at 37 °C and mixing them in a 1.5:1 molar ratio (DNA:protein). The mixture was

incubated for 50 min at this temperature, and then spun down for 5 min to remove insoluble material. Crystallization screenings were performed immediately after complex formation with a Cartesian MicroSys robot (Genomic Solutions) using the sitting-drop method with nanodrops of 0.1 µl of protein plus 0.1 µl of reservoir solution, and a reservoir volume of 60 µl. The final concentration of protein in the DNA–protein complex solution was 4 mg ml^{-1}. The best diffracting crystals were obtained under different conditions for each protein and were easily reproduced in DVX plates by hanging-drop vapour diffusion with drops of 1 µl mixed with an equal volume of reservoir solution consisting of 35% 2-ethoxyethanol in 0.1 M sodium cacodylate, pH 6.5, in the case of Amel3–Amel4–DNA–Ca^{2+} to make the complex, or 35% methanol in 0.1 M sodium cacodylate, pH 6.5, in the case of Amel3–Amel4–DNA–Mg^{2+}. The crystallization conditions for Ini3–Ini4–DNA–Ca^{2+} were 30% PEG-400, 0.1 M sodium acetate, pH 4.5, 0.2 M calcium acetate, whereas for the Ini3–Ini4–DNA–Mg^{2+} complex these conditions were 20% PEG-1000, 0.1 M imidazole, pH 8.0, 0.2 M sodium acetate. In all cases, crystals reached their definitive size in 24–48 h. Amel3–Amel4–DNA–Ca^{2+} and Ini3–Ini4–DNA–Ca^{2+} crystals were directly collected and frozen in liquid nitrogen. Amel3–Amel4–DNA–Mg^{2+} crystals were collected in 35% methanol, 0.1 M sodium cacodylate, pH 6.5, 20% ethyleneglycol, whereas Ini3–Ini4–DNA–Mg^{2+} crystals were collected in 20% PEG-1000, 0.1 M imidazole, pH 8.0, 0.2 M sodium acetate, 10% glycerol and then frozen in liquid nitrogen. The crystallization trials contained 2 mM of MgCl$_2$ or CaCl$_2$ depending of the complex that was crystallized.

Mass spectrometry. Mass determination of intact proteins was performed in a linear LTQ ion trap mass spectrometer (Thermo Finnigan) equipped with a nano-electrospray ionization source by using coated GlassTip PicoTip emitters (New Objective). Samples were desalted and concentrated with Zip Tips (Millipore) following the manufacturer's protocol. The spectrometer was operated according to the manufacturer's instructions with manual adjustment of the collision energies. Fragment spectra were interpreted manually. Nano-electrospray ionization mass spectometry analysis using coated GlassTip PicoTip emmiters (New Objective, USA) of the intact protein predominantly gave multiply protonated molecules corresponding to molecular masses of $M = 19{,}538$ Da for Amel3 and 19,983 Da for Amel4 and 19,818 Da for Ini3 and 19,709 Da for Ini4, obtained by deconvoluting the multiply charged ions (Supplementary Fig. 2) using MagTran software v.1.02 provided by Z. Zhang[30]. These experimental values are in good agreement with the theoretical values ($M_{\text{monoisotopic}} = 19{,}667, 20{,}113$ Da (Amel3–Amel4) and 19,946, 19,838 Da (Ini3–Ini4)) without the first methionine residue. To prepare the protein in the crystal for mass determination, the crystals were washed twice in a drop of the crystallization solution to eliminate any soluble non-crystalline protein and then it was dissolved in PBS. Mass determination was done with a 4800 MALDI–TOF/TOF analyser (Applied Biosystems) in linear mode using α-cyano-4-hydroxy cinnamic acid (5 mg ml^{-1} in 1:1 ACN:0.2% TFA) as matrix. The mass determination of dissolved crystals was done by MALDI–TOF/TOF instead of nano-ESI-MS owing to the presence of small amounts of PEG 1000 used in the crystallization process, which hinders nano-ESI-MS data analysis.

Data collection, structure solution, model building, refinement and structural analysis. All data were collected at cryogenic temperatures using synchrotron radiation at 100 K. Crystals were mounted and cryoprotected. The data sets were collected using synchrotron radiation at the ID23 beamlines at the ESRF, and at the PX beamline at the SLS. Diffraction data were recorded on an ADSC-Q4 or Mar225 CCD detector depending on the beamline. Processing and scaling were accomplished with HKL2000. Statistics for the crystallographic data are summarized in Supplementary Table 2. The structure was solved using the molecular replacement method as implemented in the program MOLREP or PHASER. The search model was based on a poly-alanine backbone derived from the PDB entry 1G9Z. A refined $2F_o - F_c$ map showed clear and contiguous electron density for the protein backbone and for many of the side chains. REFMAC5 was applied for refinement (Supplementary Table 2). The coordinates and structure factors have been deposited in the PDB (accession numbers 2vbj, 2vbl, 2vbn, 2vbo). The identification and analysis of the protein–DNA hydrogen bonds and van der Waals contacts was done with the Protein Interfaces, Surfaces and Assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)[31]. A summary of the hydrogen bonds, the contacting interface bases and the buried surface area is shown in Supplementary Table 3. The program NUCPLOT[32] was also used to list all the specific protein–DNA contacts discussed in the text.

***In vitro* cleavage assay conditions.** Cleavage assays were performed at 37 °C in 10 mM Tris-HCl, pH 8, 50 mM NaCl, 10 mM MgCl$_2$. Target concentration was 2 nM (XmnI linearized target substrates in plasmid pGEM-T) and protein concentrations were 120, 90, 60, 40, 30, 20, 10, 7.5, 5, 3.5, 2, 1, 0.5, 0.25 and 0 nM (lanes 1–15), or 120 nM (lane A) or 0 nM (lane B), in 25 µl final volume reaction (Fig. 1b). Reactions were stopped after 1 h by addition of 5 µl of 45% glycerol,

95 mM EDTA (pH 8), 1.5% (w/v) SDS, 1.5 mg ml⁻¹ proteinase K and 0.048% (w/v) bromophenol blue (6× buffer stop), incubated at 37 °C for 30 min and electrophoresed in a 1% agarose gel. The linearized target plasmid has 3 kb and after meganuclease cleavage yields two smaller bands of 2 kb and 1 kb. The gels were stained using SYBR Safe DNA gel staining (Invitrogen) and the intensity of the bands observed on ultraviolet light illumination was quantified with the ImageJ software (http://rsb.info.nih.gov/ij/). The percentage of cleavage was calculated with the following equation: percentage cleavage = $100 \times (I_{2\,kb} + I_{1\,kb})/(I_{3\,kb} + I_{2\,kb} + I_{1\,kb})$, where $I_{1\,kb}$, $I_{2\,kb}$ and $I_{3\,kb}$ are the intensities of the 1-, 2- or 3-kb bands, respectively.

In silico **analysis.** Results in Supplementary Tables 4 and 5 were obtained using FoldX. For the protein–DNA complex analysis (Supplementary Table 4), the difference of energies for the cross mutation of both the protein and the DNA sequences in the crystal structures were calculated. For this purpose the Amel3–Amel4 and Ini3–Ini4 amino acid sequences were modelled onto the I-CreI structure and vice versa. A similar procedure was performed with the wild-type and *XPC* DNA sequences. For the DNA analysis (Supplementary Table 5), the difference of energies when the wild-type and *XPC* DNA sequences were modelled onto the crystal structures of the DNA in the three complexes (in the absence of protein) was calculated. In Supplementary Tables 4 and 5 two types of calculations are shown: (1) for the DNA analysis, the difference of torsional energy plus van der Waals intraclashes between the mutant and the WT; (2) for the protein–DNA complex analysis, $\Delta\Delta G_{int} + \Delta$van der Waals DNA intraclashes (where $\Delta\Delta G_{int}$ is the difference in interaction energy between the mutant and the wild type). The reason to consider the increase in the intraclashes of DNA is to take into account those cases for which the energy of interaction between the protein and the DNA may be very favourable, but the mutations in the DNA make it really unstable.

Cell culture and transfection. *In vivo* experiments were done with D75N I-CreI (ref. 14). The human MRC5 cell line was cultured in MEM medium (Invitrogen Life Science) supplemented with 10% FCS (PAA), 2 mM L-glutamine, penicillin and streptomycin. CHO-K1 and CHO-π10_XPC2 cell lines were cultured in F12-K medium (Invitrogen Life Science) supplemented with 10% FCS (PAA), 2 mM L-glutamine, penicillin, streptomycin and amphotericin B. The CHO-XRCC4⁻ cell line has been described previously[33] and was cultured in DMEM medium without sodium pyruvate. Except for the γ-H2AX phosphorylation assay, all transfections were performed in a 10-cm dish with the Polyfect (Qiagen) technology. In brief, cells were seeded at 2×10^5 cells per dish one day before transfection. Meganuclease expression vectors or control vector (1 µg) were used according to the manufacturer's recommendation. For the analysis of γ-H2AX phosphorylation, cells were seeded at 10^4 cells per well in a 96-well plate and transfected with different amounts of meganuclease expression vector (as indicated in Supplementary Fig. 8), with Polyfect reagent according to the supplier's (Qiagen) protocol.

Chromosomal assay in mammalian cells. The CHO cell line (CHO-π10_XPC2) harbouring the reporter system was seeded at a density of 2×10^5 cells per 10-cm dish in complete medium (Kaighn's modified F-12 medium (F12-K)), supplemented with 2 mM L-glutamine, penicillin (100 units ml⁻¹), streptomycin (100 µg ml⁻¹), amphotericin B (Fongizone; 0.25 µg ml⁻¹; Invitrogen Life Science) and 10% FBS (Sigma-Aldrich Chimie). The next day cells were transfected with Polyfect transfection reagent (Qiagen). In brief, 2 µg of LacZ repair matrix vector was cotransfected with 1 µg of meganuclease expression vectors. After 72 h of incubation at 37 °C, cells were fixed in 0.5% glutaraldehyde at 4 °C for 10 min, washed twice in 100 mM phosphate buffer with 0.02% NP40, and stained with the following staining buffer: 10 mM phosphate buffer, 1 mM MgCl₂, 33 mM potassium hexacyanoferrate (III), 33 mM potassium hexacyanoferrate (II) and 0.1% (v/v) 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). After 16 h incubation at 37 °C, plates were examined under a light microscope and the number of LacZ-positive cell clones was counted. The frequency of LacZ repair is expressed as the number of LacZ⁺ foci divided by the number of transfected cells (5×10^5) and corrected by the transfection efficiency.

γ-H2AX immunocytochemistry in mammalian and human cells. MRC5 cells were transfected by Polyfect reagent (Qiagen) with 4 µg of DNA mixture containing 1 µg of plasmid encoding meganuclease (500 ng of each monomer) or empty vector, 200 ng Ds-Red-expressing vector and 2.8 µg of empty vector as a stuffer. Forty-eight hours after transfection, cells were fixed with 2% of paraformaldehyde for 30 min and permeabilized with 0.5% Triton for 5 min at 22 °C. After wash, cells were incubated with PBS/Triton 0.3% buffer containing 10% normal goat serum (NGS) and 3% BSA for 1 h to block non-specific staining.

CHO-K1 and CHO-π10_XPC2 cells were transfected with 1 µg of plasmid encoding meganuclease or empty vector by Polyfect reagent (Qiagen). Alternatively, cells were exposed to 2 µM of etoposide for 1 h, 2 h before harvesting. Forty-eight hours after transfection, cells were fixed with 2% of paraformaldehyde for 15 min and permeabilized with ice-cold 100% methanol for 10 min in the freezer. After wash, cells were incubated with PBS/Triton 0.3% buffer containing 10% NGS for 1 h to block non-specific staining.

Cells were then incubated with anti-γ-H2AX antibody, either for 1 h at room temperature (Upstate, 1/10,000) or overnight at 4 °C (Cell Signaling, 1/200) diluted in PBS/Triton 0.3% with 3% BSA and 10% NGS followed by 1 h incubation with the secondary antibody Alexa Fluor 488 goat anti-rabbit (Invitrogen/Molecular Probes, 1/1,000) diluted in PBS/Triton 0.3% and 3% BSA. After incubation with 1 µg ml⁻¹ 4,6-diamino-2-phenylindole (DAPI; Sigma), coverslips were mounted and analysed by fluorescent microscopy. For confocal microscopy, a Leica SP2 microscope was used to perform the immunohistochemical analysis; the microscope objective used was oil-immersed with ×40 objective with 1.2 numerical aperture. The software for acquisition was Leica LCS; the sections were acquired each micrometer, and after that the maximum projection was quantified by Metamorph (Universal Imaging Corporation).

We have previously described how a locally engineered meganuclease can cleave from one to several sequences out of the 64 10NNN and 5NNN palindromic I-CreI derivate targets, respectively[14,15]. MegaX resulted from the combination of modules with lower specificity (up to 32 and 21 10NNN and 5NNN cleaved, respectively), and should thus represent a meganuclease showing, at least in theory, a degeneracy among the highest that can be obtained by the engineering methods described previously. The first-generation zinc finger nuclease (Lʷᵗ–Rʷᵗ) has been described previously[9]. It was designed to cleave a specific target located in the human *IL2RG* gene. The heterodimer Lʷᵗ–Rʷᵗ was shown to generate a high level of γ-H2AX phosphorylation in response to DNA damage.

γ-H2AX phosphorylation in mammalian cells. γ-H2AX phosphorylation was detected by ELISA using a fluorescent assay according to the suppliers (Active Motif). XD17, a CHO-XRCC4⁻ cell line deficient in NHEJ (see Supplementary Fig. 8), and CHO-π10_XPC2 cells were transfected with 1 µg of plasmid encoding meganuclease or empty vector by Polyfect reagent (Qiagen). Forty-eight hours after the transfection or 6 h after the etoposide treatment, the cells were fixed with ice-cold 100% methanol. After washing with PBS, the cells were first incubated with the DSB marker antibody anti-phosphohistone γ-H2AX (Ser 139). After a new PBS washing, the cells were subsequently incubated with a fluorescent Chromeo 488-labelled secondary antibody. Fluorescence was further measured with an automatic 96-well plate reader as described by the suppliers (Wallac Victor 1420, Perkin Elmer).

29. Prieto, J. *et al.* The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Res.* **35**, 3262–3271 (2007).

30. Zhang, Z. & Marshall, A. G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225–233 (1998).

31. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

32. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions. *Nucleic Acids Res.* **25**, 4940–4945 (1997).

33. Delacote, F., Han, M., Stamato, T. D., Jasin, M. & Lopez, B. S. An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. *Nucleic Acids Res.* **30**, 3454–3463 (2002).

LETTERS

The essential role of the CopN protein in *Chlamydia pneumoniae* intracellular growth

Jin Huang[1], Cammie F. Lesser[1,2] & Stephen Lory[1]

Bacterial virulence determinants can be identified, according to the molecular Koch's postulates[1], if inactivation of a gene associated with a suspected virulence trait results in a loss in pathogenicity. This approach is commonly used with genetically tractable organisms. However, the current lack of tools for targeted gene disruptions in obligate intracellular microbial pathogens seriously hampers the identification of their virulence factors. Here we demonstrate an approach to studying potential virulence factors of genetically intractable organisms, such as *Chlamydia*. Heterologous expression of *Chlamydia pneumoniae* CopN in yeast and mammalian cells resulted in a cell cycle arrest, presumably owing to alterations in the microtubule cytoskeleton. A screen of a small molecule library identified two compounds that alleviated CopN-induced growth inhibition in yeast. These compounds interfered with *C. pneumoniae* replication in mammalian cells, presumably by 'knocking out' CopN function, revealing an essential role of CopN in the support of *C. pneumoniae* growth during infection. This work demonstrates the role of a specific chlamydial protein in virulence. The chemical biology approach described here can be used to identify virulence factors, and the reverse chemical genetic strategy can result in the identification of lead compounds for the development of novel therapeutics.

Chlamydia pneumoniae, a human respiratory pathogen, is associated with atherosclerosis and has been linked to heart disease and stroke[2]. This obligate intracellular pathogen resides in host cells within vacuoles referred to as inclusions[3]. *Chlamydia* usurp various host cellular processes to promote virulence[4–9], presumably through the actions of proteins that they directly secrete into host cells and/or express on the outer surface of the inclusion membrane[10–13].

The yeast *Saccharomyces cerevisiae* is an established model system that can be used to identify and characterize bacterial virulence proteins[14]. The underlying premise of this system is that many bacterial virulence proteins target cellular processes conserved from yeast to mammals. Indeed, expression of numerous bacterial virulence proteins in yeast inhibits growth owing to targeting of conserved eukaryotic cellular processes[15]. We expressed five probable *C. pneumoniae* virulence proteins in yeast. Three of these proteins, CopN, CP1062 and CP0833, are putative substrates of the *C. pneumoniae* type III system, a specialized secretion system that directly translocates proteins from the bacterial cytosol into host cells. During an infection, CopN is detected on the inclusion membrane, CP0833 in the host cell cytosol, and CP1062 at both[16]. Whereas CP0679 encodes a putative serine/threonine kinase[17], CP0358 encodes a serine/threonine protein phosphatase. As such, both encode potential virulence factors.

Expression of CopN and CP1062 severely inhibited yeast growth. This growth inhibition was alleviated when expression levels of CP1062 but not CopN were lowered (Fig. 1a). CopN inhibited yeast growth regardless of whether the protein was expressed on its own or fused to GFP (green fluorescent protein). This inhibitory activity was

also observed with expression of CopN from *Chlamydia psittaci* B577 (*Chlamydia abortus*), but not with expression of more distally related CopN homologues, including CopN of *Chlamydia trachomatis*, YopN of *Yersinia enterocolitica* and PopN of *Pseudomonas aeruginosa*.

Expression of GFP–CopN resulted in the accumulation of large-budded yeast (Fig. 1b, top panel). By 6 h post-induction of expression, 90% of the GFP–CopN-expressing yeast, but only 22% of GFP-expressing yeast (Supplementary Fig. 1), appeared as large-budded cells. Yeast normally undergo nuclear division coincident with the formation of large-budded cells, but the majority of the large-budded GFP–CopN-expressing yeast (91%) contained only a single nucleus, which was present in only one of the two buds (Fig. 1c).

To determine whether the large-budded yeast had undergone DNA replication, we quantified their DNA content using flow cytometry (FACS). Exponentially growing haploid yeast expressing GFP–CopN or GFP were synchronized at G1 and then released to progress through the cell cycle. In both cases, a predominant 2N DNA peak was
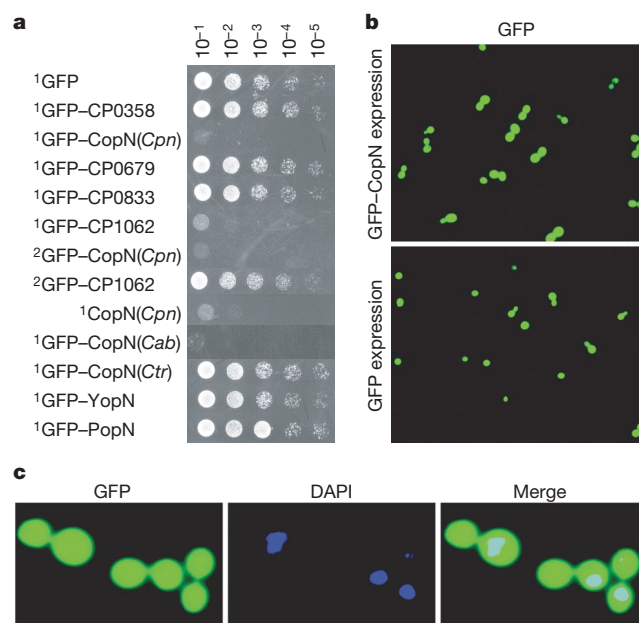

Figure 1 | CopN expression inhibits yeast growth and results in the accumulation of large-budded yeast. a, Serial dilutions of yeast that conditionally express the designated proteins were spotted on inducing media and grown for 48 h. Genes encoding the proteins were cloned on either a high (^{1}) or low (^{2}) copy number plasmid. *Cpn, C. pneumoniae*; *Cab*, *C. abortus*; *Ctr, C. trachomatis*. **b,** Image of yeast expressing GFP–CopN or GFP, visualized 6 h post-induction. **c,** Enlarged image of yeast 6 h post-induction of GFP–CopN. The yeast were fixed and stained with DAPI to visualize the nuclei.

[1]Department of Microbiology and Molecular Genetics, [2]Division of Infectious Diseases, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.
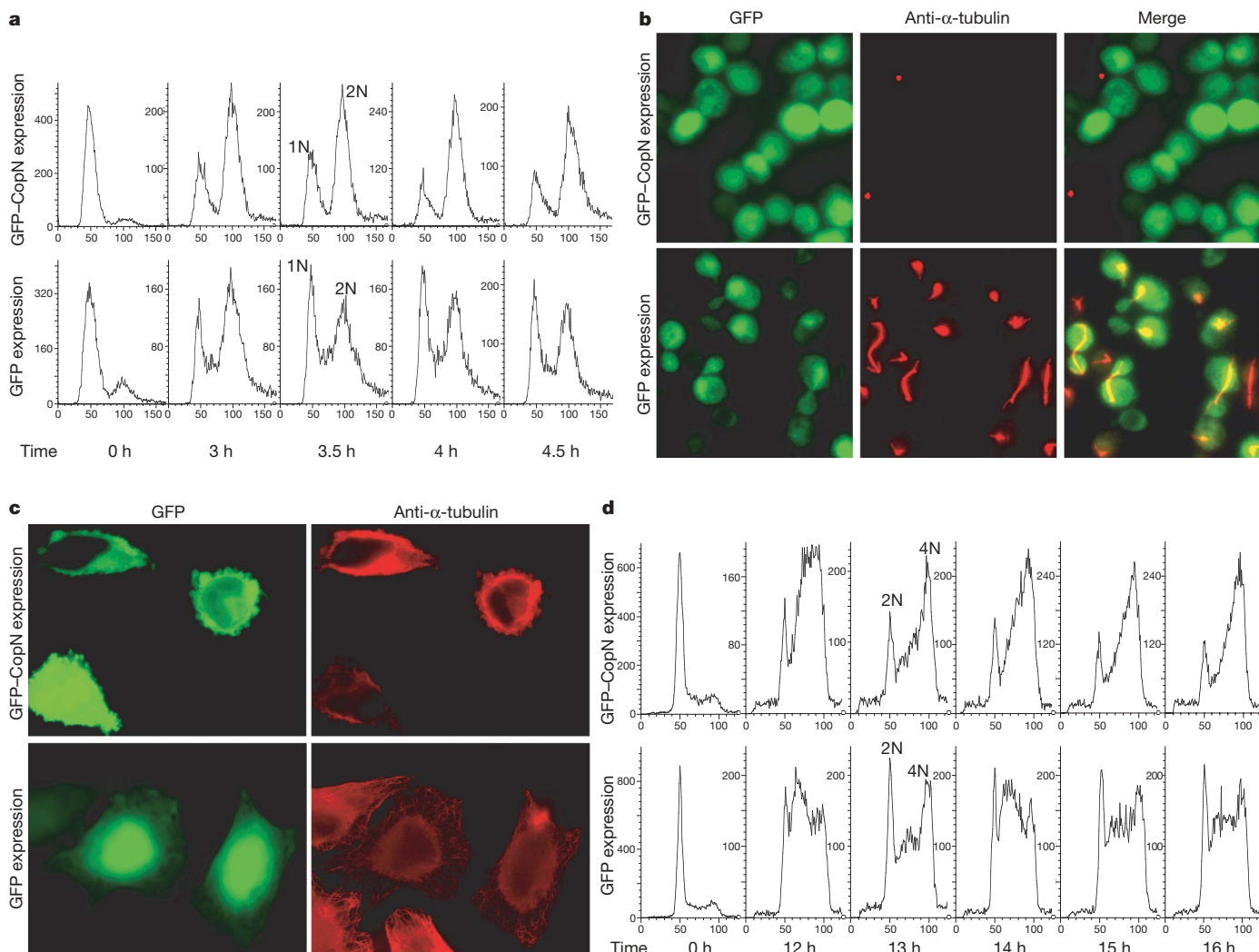
Figure 2 | CopN expression induces a cell cycle arrest in both yeast and mammalian cells due to disruption of microtubules. a, d, FACS analyses of the DNA content of **a**, yeast and **d**, 293 cells expressing GFP–CopN or GFP at the designated time points. The peaks labelled as 1N or 2N in yeast and 2N or 4N in 293 cells indicate the DNA content. **b, c,** Images of **b,** yeast 6 h post-induction of expression and **c,** HeLa cells 12 h post-transfection for transient expression of GFP–CopN or GFP. In both cases, cells were fixed and stained with anti-α-tubulin antibodies (red).

observed after 3 h, indicating that the majority of the yeast had progressed through S phase and completed DNA replication (Fig. 2a). However, while GFP-expressing yeast continued to proceed through the cell cycle, those expressing GFP–CopN arrested at this point. Thus, yeast expressing GFP–CopN arrest at the G2/M phase of the cell cycle.

Disruption of yeast microtubules can prevent formation of the spindle apparatus, which is required for mitosis, resulting in the accumulation of large-budded 2N yeast. Thus, we examined the integrity of the spindle apparatus of CopN-expressing yeast. Remarkably, no spindles were detected in GFP–CopN-expressing yeast (Fig. 2b, top panels). GFP-expressing yeast displayed normal spindles at the appropriate point in the cell cycle (Fig. 2b, bottom panels). Thus, CopN expression results in a G2/M cell cycle arrest due to disruption of spindle apparatus.

We next investigated whether the activity of CopN was conserved from yeast to mammals. We examined the structural integrity of the microtubule network in GFP–CopN-expressing epithelial cells (HeLa cells). As shown in Fig. 2c, 12 h post-transfection, microtubule networks were disrupted in GFP–CopN-expressing cells. In contrast, a characteristic radial array of microtubules was observed in GFP-expressing cells.

Disruption of the microtubules in mammalian cells can also result in a G2/M cell cycle arrest. To test whether CopN confers such a

phenotype, we established stable cell lines that conditionally express GFP–CopN or GFP. FACS analyses were performed to examine the effect of CopN expression on cell cycle progression for 16 h after release from G1 synchronization. Delay of cell cycle progression in the CopN-expressing cells was first observed at 12 h when the cells started to accumulate at the G2/M transition. This 4N peak continued to accumulate over the next four hours (Fig. 2d, top panel). In contrast, GFP-expressing cells continued to progress through the cell cycle (Fig. 2d, bottom panel). These results demonstrate that expression of CopN also induces a G2/M cell cycle arrest in mammalian cells.

Genetic tools to create *C. pneumoniae* that do not express CopN are currently unavailable. To circumvent this limitation, we screened for small molecule inhibitors of CopN activity. Specifically, we screened a library of ~40,000 small molecules for those that alleviated yeast growth inhibition due to CopN expression. Two compounds, 0433YC1 and 0433YC2 (Fig. 3a), were found to reproducibly restore growth of CopN-expressing yeast to levels 40% and 29%, respectively, of yeast expressing an inactive CopN allele (CopN R268H) (Fig. 3b). At concentrations used in the screen, these compounds did not affect growth of wild-type yeast (data not shown).

To investigate the role of CopN during a *C. pneumoniae* infection, the two inhibitors were used to essentially create 'functional knockouts' of CopN. Treatment of infected buffalo green monkey kidney
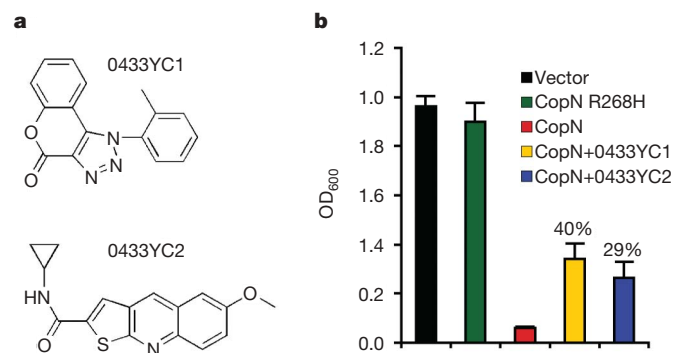
Figure 3 | The small molecule inhibitors 0433YC1 and 0433YC2 alleviate yeast growth inhibition due to CopN expression. a, Structures of compounds 0433YC1 (ChemDiv 5947-0064) and 0433YC2 (ChemDiv C303-0665). **b**, Growth of yeast (mean + s.e.m., $n = 4$) expressing either GFP, an inactive allele of GFP–CopN (R268H), or GFP–CopN in the presence and absence of 0433YC1 or 0433YC2 at 12.5 μg ml^{-1}. The percentages shown indicate the rate of restoration of growth in the presence of compounds relative to yeast expressing the inactive CopN allele. Student's t-test was performed between CopN-expressing yeast treated with 0433YC1 ($P = 0.004$) or 0433YC2 ($P = 0.02$) and untreated control.

(BGMK) cells[18] with either 0433YC1 or 0433YC2 at 10 μg ml^{-1} for 72 h resulted in a significant reduction in the replication of *C. pneumoniae* (Fig. 4a). The presence of the compounds in the media led to a decrease in *dnaK* transcription by 68–84% as compared to *dnaK* levels present in host cells grown in untreated media[19–21]. Similarly, the addition of 0433YC2 inhibited replication in Hep-2 cells (Fig. 4b). Both inhibitors interfered with the intracellular replication of

C. pneumoniae in a dose-dependent manner (Fig. 4c). No toxic effect on BGMK cells was observed when either compound was added at 20 μg ml^{-1} as assayed by either monitoring mitochondrial dehydrogenase activity or by microscopic examination of cell morphology (data not shown). Removal of 0433YC2 from the media of infected BGMK cells after 72-h treatment did not lead to an immediate recovery of *C. pneumoniae* growth (Fig. 4d). Neither of the compounds inhibited replication of *C. trachomatis* in BGMK cells (Supplementary Fig. 2). This result is perhaps not surprising, given that the expression of CopN from *C. trachomatis* did not inhibit yeast growth (Fig. 1a). Immunofluorescence microscopy revealed that the compounds also inhibited the development of *C. pneumoniae* inclusions observed within the infected BGMK cells (Fig. 4e, f and Supplementary Fig. 3). Infected cells treated with the compounds essentially lacked large inclusions (Fig. 4e, f) characteristic of *C. pneumoniae* growth in host cells (2.3–5 μm in diameter) (Fig. 4g). Rather, small inclusions were observed in cells incubated with the more potent compound 0433YC2 (8.8 inclusions per cell and 0.4–0.6 μm in diameter). These small inclusions resembled those seen in cells treated with chloramphenicol (9.2 inclusions per cell and 0.4–0.6 μm; Fig. 4h), an antimicrobial agent active against *Chlamydia*.

Taken together, our data demonstrate that CopN is required to support the intracellular growth of *C. pneumoniae* and plays an essential virulence role in a cell culture model of infection[22]. By using the CopN small molecule inhibitors identified in yeast, we were able to fulfil the molecular Koch's postulates to identify the first chlamydial protein required for virulence of this obligate intracellular organism. This strategy can be extended to study candidate virulence factors from other pathogens, especially those, like *Chlamydia* species, that are genetically intractable. For example, more than 30 *C. trachomatis* candidate virulence proteins inhibit yeast growth[23].
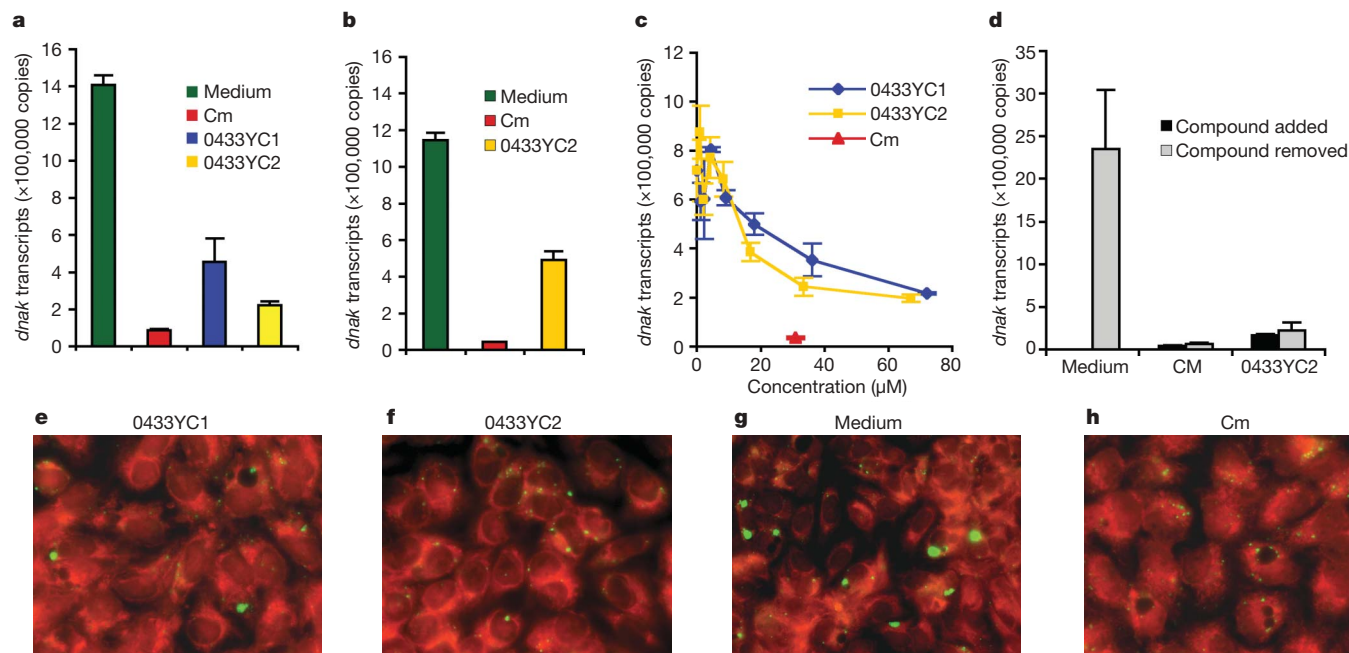

Figure 4 | The CopN inhibitors 0433YC1 and 0433YC2 inhibit *C. pneumoniae* replication in host cells. a–d, Growth of *C. pneumoniae* in host cells monitored by quantification of *dnaK* transcription by real-time PCR. Host cells were infected for one hour at a multiplicity of infection (m.o.i.) of 10:1 followed by incubation in fresh media containing the inhibitors. Each of the assays was repeated 3–10 times. Student's *t*-test was performed between treated and untreated cells. **a**, BGMK cells treated with compounds 0433YC1 ($P = 0.0004$) and 0433YC2 ($P = 0.000001$) at 10 μg ml^{-1}. Standard growth medium and that plus chloramphenicol (Cm) were used as growth and inhibition controls. **b**, Hep-2 cells treated with 0433YC2 ($P = 0.0016$) at 10 μg ml^{-1}. **c**, BGMK cells treated with compounds at the designated concentrations (μM) on the *x* axis. **d**, BGMK cells were first treated with chloramphenicol or 0433YC2 for 72 h. The compounds were then removed and *dnaK* levels were determined after an additional 48 h. Data in **a**–**d** are presented as total copy number of *dnaK* transcripts per well of the 24-well plate (mean + s.e.m.). **e**–**h**, Immunofluorescent images of BGMK cells infected with *C. pneumoniae* at an m.o.i. of 10:1. The *C. pneumoniae* inclusions are stained with anti-*Chlamydia*-LPS antibody (green) and the host cell is counterstained red. Cells were treated with **e**, 0433YC1, **f**, 0433YC2, **g**, media control, or **h**, chloramphenicol.

The expression of candidate virulence proteins in yeast can also result in new insights into the roles of these proteins in pathogenesis[15]. Our observation that heterologous expression of CopN in both yeast and mammalian cells affected the formation of microtubule structures and caused a cell cycle phase-specific cell division block is intriguing, and suggests that CopN directly or indirectly targets microtubules during the course of an infection. Infection with *Chlamydia* has been observed to delay host cell division when high titres of *Chlamydia* are found in the host cells[4,24]. Thus, if CopN does induce a cell cycle block, this could potentially divert resources of the infected cell to favour the multiplication of *Chlamydia*, a strategy used by other bacterial pathogens to facilitate infection[25,26].

CopN is a member of a family of proteins common to pathogenic organisms including *C. psittaci* B577 (*C. abortus*), *C. trachomatis*, *Yersinia* species and *P. aeruginosa*. Expression of only CopN from *C. pneumonia* and its closest homologue from *C. psittaci* (Supplementary Fig. 4) inhibited yeast growth (Fig. 1a). *Yersinia* YopN has been implicated to regulate type III secretion by controlling access to the secretory channel. Following contact of the bacteria with a eukaryotic cell, YopN is translocated along with other effectors into the cytoplasm of the host cell[27,28]. Our results indicate that the limited homology of CopN to YopN may account for the lack of yeast toxicity of YopN and that CopN may play multiple roles in pathogenesis, including regulation of secretion and modification of the host microtubule network. There is a precedent for this phenomenon, as *Shigella* IpaB and *Salmonella* SipB, both components of the type III secretion system, are also delivered into host cells where they interact with caspase-1 to trigger apoptosis. In contrast, their *Yersinia* homologue, YopB, a component of the translocon, is not known to target caspase-1 (refs 29, 30).

Interestingly, the small molecule inhibitors of CopN that inhibited infection with *C. pneumoniae* did not inhibit infection with *C. trachomatis*, implying that the small molecule inhibitors target a function of CopN not shared among its more distal relatives. Thus, the specificity of our compounds for *C. pneumoniae* suggests that we have identified lead compounds for the development of therapeutics that specifically target *C. pneumoniae* and its closer relatives.

METHODS SUMMARY

The effect of expression of chlamydial proteins on yeast (W303) was determined by measuring growth 48 h after the cultures were spotted onto solid inducing media. Cellular activities of CopN were examined using immunofluorescence microscopy and flow cytometry (FACS). The high-throughput yeast growth suppression screen with a 40,000 compound chemical library (Supplementary Table 1) used RDY0433 expressing GFP–CopN. RDY0433 is a yeast strain that does not express PDR1 and PDR3, two major drug efflux pumps. The effect of compounds on growth of *C. pneumoniae* following infection of BGMK cells with *C. pneumoniae* strain AR39 was determined by immunofluorescence staining and by measurements of copies of *dnaK* transcripts using RT–PCR. Statistical analysis was performed with Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1.   Falkow, S. Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* **10** (Suppl. 2), S274–S276 (1988).
2.   Campbell, L. A. & Kuo, C. C. *Chlamydia pneumoniae* — an infectious risk factor for atherosclerosis? *Nature Rev. Microbiol.* **2**, 23–32 (2004).
3.   Hackstadt, T. in *Chlamydia: Intracellular Biology, Pathogenesis, and Immunity* (ed. Stephens, R. S.) 101–138 (ASM Press, 1999).
4.   Crocker, T. T., Pelc, S. R., Nielsen, B. I., Eastwood, J. M. & Banks, J. Population dynamics and deoxyribonucleic acid synthesis in HeLa cells infected with an ornithosis agent. *J. Infect. Dis.* **115**, 105–122 (1965).
5.   Hackstadt, T., Scidmore, M. A. & Rockey, D. D. Lipid metabolism in *Chlamydia trachomatis*-infected cells: Directed trafficking of Golgi-derived sphingolipids to the chlamydial inclusion. *Proc. Natl Acad. Sci. USA* **92**, 4877–4881 (1995).
6.   Fan, T. *et al.* Inhibition of apoptosis in *Chlamydia*-infected cells: Blockade of mitochondrial cytochrome c release and caspase activation. *J. Exp. Med.* **187**, 487–496 (1998).
7.   Carabeo, R. A., Grieshaber, S. S., Fischer, E. & Hackstadt, T. *Chlamydia trachomatis* induces remodeling of the actin cytoskeleton during attachment and entry into HeLa cells. *Infect. Immun.* **70**, 3793–3803 (2002).
8.   Carabeo, R. A., Mead, D. J. & Hackstadt, T. Golgi-dependent transport of cholesterol to the *Chlamydia trachomatis* inclusion. *Proc. Natl Acad. Sci. USA* **100**, 6771–6776 (2003).
9.   Su, H. *et al.* Activation of Raf/MEK/ERK/cPLA2 signaling pathway is essential for chlamydial acquisition of host glycerophospholipids. *J. Biol. Chem.* **279**, 9409–9416 (2004).
10. Hsia, R. C., Pannekoek, Y., Ingerowski, E. & Bavoil, P. M. Type III secretion genes identify a putative virulence locus of *Chlamydia. Mol. Microbiol.* **25**, 351–359 (1997).
11. Fields, K. A., Mead, D. J., Dooley, C. A. & Hackstadt, T. *Chlamydia trachomatis* type III secretion: Evidence for a functional apparatus during early-cycle development. *Mol. Microbiol.* **48**, 671–683 (2003).
12. Peters, J., Wilson, D. P., Myers, G., Timms, P. & Bavoil, P. M. Type III secretion a la *Chlamydia. Trends Microbiol.* **15**, 241–251 (2007).
13. Valdivia, R. H. *Chlamydia* effector proteins and new insights into chlamydial cellular microbiology. *Curr. Opin. Microbiol.* **11**, 53–59 (2008).
14. Lesser, C. F. & Miller, S. I. Expression of microbial virulence proteins in *Saccharomyces cerevisiae* models mammalian infection. *EMBO J.* **20**, 1840–1849 (2001).
15. Siggers, K. A. & Lesser, C. F. The yeast *Saccharomyces cerevisiae*: A versatile model system for the identification and characterization of bacterial virulence proteins. *Cell Host Microbe* **4**, 8–15 (2008).
16. Lugert, R., Kuhns, M., Polch, T. & Gross, U. Expression and localization of type III secretion-related proteins of *Chlamydia pneumoniae. Med. Microbiol. Immunol. (Berl.)* **193**, 163–171 (2004).
17. Verma, A. & Maurelli, A. T. Identification of two eukaryote-like serine/threonine kinases encoded by *Chlamydia trachomatis* serovar L2 and characterization of interacting partners of Pkn1. *Infect. Immun.* **71**, 5772–5784 (2003).
18. Li, D. *et al.* High-yield culture and purification of Chlamydiaceae bacteria. *J. Microbiol. Methods* **61**, 17–24 (2005).
19. Khan, M. A., Potter, C. W. & Sharrard, R. M. A reverse transcriptase-PCR based assay for in-vitro antibiotic susceptibility testing of *Chlamydia pneumoniae. J. Antimicrob. Chemother.* **37**, 677–685 (1996).
20. Cross, N. A. *et al.* Antimicrobial susceptibility testing of *Chlamydia trachomatis* using a reverse transcriptase PCR-based method. *Antimicrob. Agents Chemother.* **43**, 2311–2313 (1999).
21. Huang, J. *et al.* The quantity of nitric oxide released by macrophages regulates *Chlamydia*-induced disease. *Proc. Natl Acad. Sci. USA* **99**, 3914–3919 (2002).
22. Moulder, J. W. The relation of basic biology to pathogenic potential in the genus *Chlamydia. Infection* **10** (Suppl 1), S10–S18 (1982).
23. Sisko, J. L., Spaeth, K., Kumar, Y. & Valdivia, R. H. Multifunctional analysis of *Chlamydia*-specific genes in a yeast expression system. *Mol. Microbiol.* **60**, 51–66 (2006).
24. Horoschak, K. D. & Moulder, J. W. Division of single host cells after infection with chlamydiae. *Infect. Immun.* **19**, 281–286 (1978).
25. Oswald, E., Nougayrede, J. P., Taieb, F. & Sugai, M. Bacterial toxins that modulate host cell-cycle progression. *Curr. Opin. Microbiol.* **8**, 83–91 (2005).
26. Shafikhani, S. H. & Engel, J. *Pseudomonas aeruginosa* type III-secreted toxin ExoT inhibits host-cell division by targeting cytokinesis at multiple steps. *Proc. Natl Acad. Sci. USA* **103**, 15605–15610 (2006).
27. Lee, V. T., Anderson, D. M. & Schneewind, O. Targeting of *Yersinia* Yop proteins into the cytosol of HeLa cells: One-step translocation of YopE across bacterial and eukaryotic membranes is dependent on SycE chaperone. *Mol. Microbiol.* **28**, 593–601 (1998).
28. Day, J. B., Ferracci, F. & Plano, G. V. Translocation of YopE and YopN into eukaryotic cells by *Yersinia pestis* yopN, tyeA, sycN, yscB and lcrG deletion mutants measured using a phosphorylatable peptide tag and phosphospecific antibodies. *Mol. Microbiol.* **47**, 807–823 (2003).
29. Chen, Y., Smith, M. R., Thirumalai, K. & Zychlinsky, A. A bacterial invasin induces macrophage apoptosis by binding directly to ICE. *EMBO J.* **15**, 3853–3860 (1996).
30. Guichon, A., Hersh, D., Smith, M. R. & Zychlinsky, A. Structure-function analysis of the *Shigella* virulence factor IpaB. *J. Bacteriol.* **183**, 1269–1276 (2001).

METHODS

Plasmids and expression constructs. Original plasmid vectors and derived expression constructs are summarized in Supplementary Table 2. For yeast expression, the open reading frames of the *C. pneumoniae* genes (CP0358, CP0433, CP0679, CP0833 and CP1062) were PCR amplified from *C. pneumoniae* AR39 chromosomal DNA prepared as described[21], and cloned by the Gateway technology (Invitrogen) into the yeast high-copy plasmid pDSTY1, a gateway-adapted 2μ-based pFUS[14], which created expression constructs pY1(CP0358), pY1(CP0433), pY1(CP0679), pY1(CP0833) and pY1(CP1062). The same strategy was used for the cloning of CopN from *C. trachomatis* L2 genomic DNA (provided by Z. R. Balsara and M. N. Starnbach at Harvard Medical School), YopN from *Y. entercolitica* pYVe227 plasmid DNA (provided by V. T. Lee currently at University of Maryland), and PopN from *P. aeruginosa* PAO1 genomic DNA. This cloning allows for generation of N-terminal GFP fusion proteins under the control of the GAL10 promoter. The fragments containing the GAL10 promoter, GFP fusion gene and the ADH terminator from pFUS, pY1(CP0433) and pY1(CP1062) constructs were subcloned into the centromere-based (cen) pRS313 (refs 14, 31) through homologous recombination-mediated DNA replacement to make low-copy versions of GAL10–GFP–CopN and GAL10–GFP–CP1062 expression constructs pRS(0433) and pRS(1062). Integrating versions of the of GAL10–GFP (vector control) and GAL10–GFP–CopN were made by deleting the 2μ replication origin from the backbone of pFUS and pY1(0433) constructs which then gave rise to pYGFP/int and pY0433/int to target integration at the yeast chromosomal LEU2 locus. The high-copy (2μ) plasmid vector pDSTY3 is the non-GFP version of pDSTY1 modified by deleting the GFP open reading frame, and was used to create pY3(0433) construct for expressing pure CopN protein in the yeast. For transient mammalian expression, the CP0433 open reading frame was cloned by the Gateway technology into the vector pDEST53 (Invitrogen) to create pM53(CP0433) where expression of the GFP–CopN fusion protein is driven from a constitutive CMV promoter. The GFP expression construct pM53(GFP) was made by the removal of the *att* cassette containing the chloramphenicol resistance gene and *ccd*B gene from pDEST53 following restriction digestion with NotI and PacI, blunt-end treatment, and self-ligation. For integration and 'stably' regulated mammalian expression, the genes for GFP and GFP–CopN fusion protein were PCR amplified from the pM53(CP0433), and inserted into the EcoRV and NotI sites of the pcDNA5/FRT/TO (Invitrogen) to create pM5to(GFP) and pM5to(0433) for targeted chromosomal integration; thereafter expression of the GFP or GFP–CopN is regulated by a tetracycline-inducible CMV promoter. The pOG44 vector (Invitrogen) was used for the Flp recombinase expression in mammalian cells.

Random mutagenesis. Mutagenesis of the gene CP0433 was carried out with the GeneMorph II Random Mutagenesis kit following the manufacture's instruction (Stratagene). The target DNA was pY1(0433) and primers were the universal *attB* primers (Invitrogen). The pool of mutagenized PCR products and the BglII and BsiWI (within the CP0433 insert)-linearized pY1(0433) were used to co-transform the yeast wherein the mutagenized CP0433 gene fragments were incorporated into the yeast expression vector pDSTY1 through *in vivo* homologous recombination and gap repair. The growth of resulting yeast transformants was selected on inducing selective medium plate supplemented with 2% galactose, and the plasmids were recovered for sequencing analysis.

Yeast strains and growth assay. The yeast strains for episomal and integrative expression of bacterial genes were created by transformation of the yeast strain W303a using different plasmid expression constructs (Supplementary Table 2) and a lithium acetate method[32]. Yeast growth assays were conducted and the yeast growth rates of individual strains were compared as described[14]. Briefly, saturated overnight cultures of the strains of interest were grown in non-inducing selective synthetic media supplemented with 2% raffinose. Each culture was normalized to $OD_{600} = 1$ and then serial tenfold dilutions (5 μl) were spotted onto inducing media. The plates were incubated at 30 °C and photographs of the plates were taken 48 h after plating.

Yeast microscopy and immunofluorescence. Yeast strains carrying the plasmid constructs of interest were grown overnight in non-inducing selective synthetic media supplemented with 2% raffinose. Yeast cells were diluted to $OD_{600} = 0.5$–0.6 and grown for an additional 1 h. Then 2% galactose was added to induce expression of the fusion protein. For examination of budding morphology and nuclear division, yeast cells sampled at the designated time points were re-suspended in mounting media containing DAPI (Sigma). For immunofluorescence observation of microtubules, yeast cells were fixed in 3.7% formaldehyde, and stained with rat anti-α-tubulin antibody YOL1/34 (SeroTec) and the secondary antibody Texas Red dye-conjugated donkey anti-rat IgG (H+L, Jackson ImmunoResearch Laboratory) followed by DAPI (Sigma) staining of DNA as described[33]. All microscopic observations were performed on an

inverted Nikon Eclipse TE2000-U microscope. Images were generated by using MetaMorph software, converted to .tif format, and then transferred into Adobe Photoshop CS2 Version 9.0.2 (Adobe Microsystems) where they were re-sized, contrast-enhanced, pseudocoloured, and/or merged.

Yeast cell synchronization and flow cytometry. The yeast strains Y0433 and YGFP, derivatives of W303a carrying integrating versions of the GFP–CP0433 fusion gene and the GFP gene for integrative expression of the GFP–CopN fusion protein and GFP, were grown to early-log phase in non-inducing selective synthetic media supplemented with 2% raffinose at 30 °C. Then α-factor (10 μg ml⁻¹, Zymo Research) was added to synchronize the yeast cells at G1 phase for a total of 3 h as described[34]. At 1.5 h after addition of the α-factor, 2% galactose was added to induce protein expression. The yeast cells were released from α-factor-arrest 1.5 h after the galactose was added to the media where designated as '0 h' point, and continued to grow in inducing selective synthetic media supplemented with 2% galactose at 30 °C. Yeast were harvested at 30 min intervals and fixed in 70% ethanol. For flow cytometric analysis of DNA contents, DNA of the fixed yeast cells was stained using the fluorescence marker propidium iodide (PI, Sigma) as previously described[35]. Samples were analysed on the FACScan flow cytometer using ModFit software. The distribution of fluorescence intensity from individual cells is presented as histograms, with the *x* axis showing fluorescence intensity whereas the *y* axis shows cell number.

Mammalian cell transfection and immunofluorescence. HeLa cells were grown in DMEM medium supplemented with 10% fetal bovine serum (Invitrogen) in an incubator at 37 °C, 5% CO_2. Chemical transfection of HeLa cells was performed with GeneJuice transfection reagent according to the manufacturer's instruction (Novagen). Transfected HeLa cells were cultured for 12 h. All immunostaining procedures were performed at room temperature according to the online protocol of the Mitchison laboratory authored by A. Desai at Harvard Medical School (mitchison.med.harvard.edu/protocols). Cells were rinsed in BRB80 (80 mM PIPES, 1 mM EGTA and 1 mM MgCl₂, pH 6.8) and fixed for 10 min in 0.5% glutaradehyde in BRB80. Cell membranes were permeabilized for 15 min with a solution of 1% Triton X-100 in PBS (12 mM phosphate, 137 mM NaCl and 3 mM KCl, pH 7.4). Free aldehydes were quenched three times with NaBH₄ (1 mg ml⁻¹, Sigma) in PBS for 10 min each. Fixed cells were rinsed three times with PBST (PBS + 0.1% Triton X-100) and blocked in 1% bovine serum albumin (BSA) in PBST for 20 min. All subsequent rinses between antibody incubations were performed using PBST. All antibodies were diluted in 1% BSA in PBST. For immunofluorescence of microtubules, cells were incubated for 60 min in 1/8,000 mouse anti-α-tubulin primary antibody (B-5-1-2, Sigma) followed by a 60-min incubation in 1/2,000 Alexa Fluor 594-conjugated goat anti-mouse IgG (H+L) secondary antibody (Invitrogen). For labelling DNA, cells were incubated in DAPI (10 mg ml⁻¹, Sigma) for 20 min. After the coverslips had been washed three times with PBS and once with deionized water, they were mounted and observed on an inverted Nikon Eclipse TE2000-U microscope. Images were generated by using MetaMorph software, converted to .tif format, and then transferred into Adobe Photoshop CS2 Version 9.0.2 (Adobe Microsystems) where they were re-sized, contrast-enhanced, pseudocoloured and/or overlaid.

Stable cell lines and mammalian cell flow cytometry. Using Lipofectamine 2000 (Invitrogen), the Flp-In T-REx 293 cells (Invitrogen) were co-transfected with the plasmid construct pM5to(0433) or pM5to(GFP) along with the Flp recombinase-expressing plasmid pOG44 according to the manufacturer's instructions (Invitrogen). Being selected in the presence of hygromycin (100 μg ml⁻¹) and blasticidin (15 μg ml⁻¹), individual colonies were tested for tetracycline (1 μg ml⁻¹)-regulatable expression of the relevant constructs by fluorescence microscopy and western blotting. The resulting stable cell lines TR293-CopN and TR293-GFP capable of expressing CopN and GFP, respectively, were maintained in the constant selection of hygromycin (50 μg ml⁻¹) and blasticidin (15 μg ml⁻¹). For flow cytometric analysis of DNA contents, (5–10) × 10⁵ cells were seeded in individual T-25 cell culture flasks. Following a 24-h incubation, the semi-confluent cells were synchronized in the G1 stage with amphidicolin (5 μg ml⁻¹) for a total of 20 h. Removal of amphidicolin defined the '0 h' time point. To ensure the presence of the protein at work immediately after release of the G1 synchronization, tetracycline (1 μg ml⁻¹) was added to the cultures to initiate the protein expression at the −12 h point, and the post-synchronization induction continued for 18 h. Cells in one flask from each different cell line were collected at 1-h intervals starting at the 0-h point, and then fixed and permeabilized in 4 ml of 75% ethanol in PBS at −20 °C for at least 16 h. After one wash with PBS containing 1% BSA, the cell pellets were resuspended in 1 ml of solution containing propidium iodide (50 μg ml⁻¹; Sigma), RNase (240 μg ml⁻¹; Sigma) and Triton X-100 (0.01% v/v; Sigma). Cells were stained for at least 30 min in the dark before cell cycle analysis. The distribution

of cells in the various phases of the cell cycle was analysed on a Becton-Dickinson FACScan flow cytometer using ModFit software.

Chemical library and high-throughput screen. The screening was performed at the Institute of Chemistry and Cell Biology (ICCB) at Harvard Medical School. The isogenic yeast strain RDY0433 capable of integratively expressing GFP–CopN was screened in the CopN-based yeast growth interference assay against a pilot library of 40,000 small-molecule compounds representing a diverse portion of the ICCB collection from multiple sources (Supplementary Table 1). The assay strains were constructed from the drug-sensitive strain RDY84 (MAT**a**, pdr1DKAN, pdr3DHIS5+, ade2, trp1, his3, leu2, ura3, can1), a derivative of *S. cerevisiae* W303a lacking the major efflux pumps PDR1 and PDR3 (ref. 36). The screen assay was validated[37] by a genetically introduced point mutation to create the mutant CopN R268H that completely eliminated the inhibitory growth effect of the wild type CopN. All primary screening was done in duplicate in 384-well plates (Costar, Corning). DMSO stock compounds were transferred using pin arrays from a stock solution of $5\,\mathrm{mg\,ml^{-1}}$ to the wells, each of which was pre-filled with $30\,\mu l$ of inducing synthetic selective media containing 2% galactose. Then $10\,\mu l$ of RDY0433 cells diluted to an OD_{600} of 0.16 in inducing synthetic media containing 2% galactose were added immediately to 384-well plates. The final volume in each well was $40\,\mu l$, which contained DMSO at a final concentration of 2% and compounds at a final concentration of about $12.5\,\mathrm{mg\,ml^{-1}}$. As a positive control, cells of the strain RDY0433(R268H) carrying integrated CopN R268H were similarly constructed, grown and diluted to the same OD_{600} and then inoculated. The negative control was the assay strain RDY0433 without compounds. All plates were incubated at $30\,^{\circ}C$ for 40–42 h, and OD_{600} was read with a microtitre plate reader (Molecular Device). The effect of compounds was measured as a percentage of growth restoration using the following equation: percentage of growth restoration = $[(OD_t - OD_n)/(OD_p - OD_n)] \times 100$, where OD_t is OD_{600} of the well with the assay strain RDY0433 and test compounds, OD_n is the median value of OD_{600} of the RDY0433 cells without compounds, and OD_p is the median value of OD_{600} of RDY0433(R268H) cells without compounds. Compounds showing $\geq 10\%$ of growth restoration in duplicate tests were scored as hits[38]. Hit compounds identified from the primary screening were confirmed by repeating the growth restoration assay in 96-well plates (Costar, Corning). The compounds were tested against other isogenic strains expressing different proteins that also elicit lethal phenotypes but are not related to CopN.

C. pneumoniae infection and immunofluorescence. *C. pneumoniae* strain AR39 (53592; ATCC) was cultured in BGMK cells and the inclusion forming units of partially purified elementary bodies (EBs) were determined as previously described[21]. Test of small molecule compounds on *C. pneumoniae* growth was performed in the BGMK cell culture in 24-well cell culture plates (Costar, Corning) in 5% CO_2 at $37\,^{\circ}C$, each well containing 1 ml of growth medium. The confluent monolayer BGMK cells were infected with EBs at a multiplicity of infection (m.o.i.) of 10 by centrifugation at $35\,^{\circ}C$ with $1,200g$ for 1 h, washed twice with Hanks' balanced salt solution, and incubated in the fresh medium plus 0.2% DMSO, with or without compounds, for up to 72 h. Compounds were all used at a final concentration of $10\,\mu g\,ml^{-1}$ except in the dose-dependence experiment where compounds were used at 0.3125, 0.625, 1.25, 2.5, 5, 10 and $20\,\mu g\,ml^{-1}$ corresponding to 1.125, 2.25, 4.5, 9, 18, 36 and 72.1 μM for 0433YC1, and 1.0, 2.1, 4.2, 8.4, 16.8, 33.5 and 67 μM for 0433YC2. Chloramphenicol ($10\,\mu g\,ml^{-1}$)-treated and untreated BGMK cells were also prepared as the positive inhibition (no-growth) and negative inhibition (growth) controls, respectively. Cells from three or four wells as indicated for each concentration of compounds and from the positive and negative control wells were harvested for RNA extraction and subsequent RT–PCR. For visualization of chlamydial inclusions, the chlamydial infection and compound treatments were carried out following the same procedures in the BGMK cells cultured on coverslips in wells of 24-well plates. After incubation for 72 h, cells were washed with PBS, fixed with 100% methanol, and stained with FITC-
conjugated mAb against chlamydial LPS of the Chlamydia Culture Confirmation System (Pathfinder, BIO-RAD). Hep-2 cells were grown and treated with the same procedures as used for BGMK cells.

RNA extraction and real-time RT–PCR. Total RNA was extracted from single inoculated wells by using the RNAqueous-Micro kit (Ambion) in accordance with the manufacturer's instructions. The extracted RNAs were treated with DNase I included in the kit to eliminate the contaminating DNA. The DNA-free RNAs were confirmed by PCR without RT. RT was performed using the reverse primer specific for *C. pneumoniae dnaK* gene with the SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. The resulting cDNAs were then subjected to the real-time PCR with primers specific for *C. pneumoniae dnaK* gene and with the use of the Platinum SYBR Green qPCR SuperMix-UDG kit (Invitrogen) following the manufacturer's instructions on the ABI PRISM 7700 Sequence Detection System.

31. Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119–122 (1995).
32. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).
33. Miller, R. K. Monitoring spindle assembly and disassembly in yeast by indirect immunofluorescence. *Methods Mol. Biol.* **241**, 341–352 (2004).
34. Day, A., Schneider, C. & Schneider, B. L. Yeast cell synchronization. *Methods Mol. Biol.* **241**, 55–76 (2004).
35. Zhang, H. & Siede, W. Analysis of the budding yeast *Saccharomyces cerevisiae* cell cycle by morphological criteria and flow cytometry. *Methods Mol. Biol.* **241**, 77–91 (2004).
36. Dorer, R. K. *et al.* A small-molecule inhibitor of Mps1 blocks the spindle-checkpoint response to a lack of tension on mitotic chromosomes. *Curr. Biol.* **15**, 1070–1076 (2005).
37. Tugendreich, S. *et al.* A streamlined process to phenotypically profile heterologous proteins in parallel using yeast cell-based assays. *Genome Res.* **11**, 1899–1912 (2001).
38. Perkins, E. *et al.* Novel inhibitors of poly(ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. *Cancer Res.* **61**, 4175–4183 (2001).

LETTERS

The CRAC channel consists of a tetramer formed by Stim-induced dimerization of Orai dimers

Aubin Penna[1], Angelo Demuro[2], Andriy V. Yeromin[1], Shenyuan L. Zhang[1], Olga Safrina[1], Ian Parker[1,2] & Michael D. Cahalan[1,3]

Ca^{2+}-release-activated Ca^{2+} (CRAC) channels underlie sustained Ca^{2+} signalling in lymphocytes and numerous other cells after Ca^{2+} liberation from the endoplasmic reticulum (ER). RNA interference screening approaches identified two proteins, Stim[1,2] and Orai[3–5], that together form the molecular basis for CRAC channel activity[6,7]. Stim senses depletion of the ER Ca^{2+} store and physically relays this information by translocating from the ER to junctions adjacent to the plasma membrane[1,8,9], and Orai embodies the pore of the plasma membrane calcium channel[10–12]. A close interaction between Stim and Orai, identified by co-immunoprecipitation[12] and by Förster resonance energy transfer[13], is involved in the opening of the Ca^{2+} channel formed by Orai subunits. Most ion channels are multimers of pore-forming subunits surrounding a central channel, which are preassembled in the ER and transported in their final stoichiometry to the plasma membrane. Here we show, by biochemical analysis after cross-linking in cell lysates and intact cells and by using non-denaturing gel electrophoresis without cross-linking, that Orai is predominantly a dimer in the plasma membrane under resting conditions. Moreover, single-molecule imaging of green fluorescent protein (GFP)-tagged Orai expressed in *Xenopus* oocytes showed predominantly two-step photobleaching, again consistent with a dimeric basal state. In contrast, co-expression of GFP-tagged Orai with the carboxy terminus of Stim as a cytosolic protein to activate the Orai channel without inducing Ca^{2+} store depletion or clustering of Orai into punctae yielded mostly four-step photobleaching, consistent with a tetrameric stoichiometry of the active Orai channel. Interaction with the C terminus of Stim thus induces Orai dimers to dimerize, forming tetramers that constitute the Ca^{2+}-selective pore. This represents a new mechanism in which assembly and activation of the functional ion channel are mediated by the same triggering molecule.

As an atypical four transmembrane-spanning protein with no sequence similarity to any other ion channel, the subunit organization of Orai and the mode of activation remain undefined. Knowledge of Orai stoichiometry is crucial to understanding the mechanisms of channel assembly, gating and ion permeation, but previous studies have led to differing conclusions, with biochemical experiments suggesting that the Orai complex is a dimer in both resting and thapsigargin-treated cells[14], whereas functional measurements of expressed tandem Orai multimers indicate a tetramer as the active CRAC channel pore[15]. To resolve this issue, we began by applying biochemical techniques used in the past to solve the pore stoichiometry of other channels[16–19], including: co-immunoprecipitation, chemical cross-linking of intact and lysed cells, and polyacrylamide gel electrophoresis (PAGE) under non-dissociating conditions. Reciprocal co-immunoprecipitation confirmed that Orai subunits bearing different tags can co-assemble[11,14] (Supplementary Fig. 1). After treatment of

total cell lysates from *Drosophila* S2 cells transfected with haemagglutinin (HA)-tagged Orai, increasing concentrations of three different lysine-reactive homobifunctional reagents produced cross-linked species on SDS–PAGE gels of ~80, ~120 and ~160 kDa, consistent with molecular masses of Orai dimers, trimers and tetramers, respectively (Fig. 1a). The bands corresponding to Orai dimers were invariably the most intense. This pattern was also seen in intact Orai-transfected S2 cells using both cysteine- and lysine-reactive homobifunctional cross-linkers, again suggesting that Orai dimers are the main form of the protein present in living cells (Supplementary Fig. 2a). The relative mobility of each band decreased as a logarithmic function of the estimated number of cross-linked subunits, indicating that the cross-linked products are integral homomultimers of the monomeric subunit[18] (Supplementary Fig. 2b). This was confirmed using a functional GFP–Flag-tagged Orai construct (GFP–Orai); the GFP tag increased the apparent molecular mass of the Orai monomer by the predicted amount of ~26 kDa (Supplementary Fig. 2c, d). If each oligomeric species was composed purely of Orai, their sizes would be directly proportional to an integer multiple of the ~68 kDa GFP–Orai monomer, which is exactly what we observed. The absence of graded formation of oligomers with a higher order than dimers as a function of increasing cross-linker concentration (Fig. 1a and Supplementary Fig. 2a) or time of cross-linker incubation (Supplementary Fig. 3) suggests that the Orai oligomeric state is a dimer in resting S2 cells. The predominant homodimeric stoichiometry was further confirmed in S2 and human embryonic kidney HEK293 cells by perfluorooctanoic acid (PFO) native gel electrophoresis. PFO is a mild non-denaturing and non-dissociating detergent for molecular complexes that has been successfully used to determine the quaternary structure of various membrane proteins[19]. Under mild solubilization conditions, Orai was almost exclusively observed in a dimeric state (Fig. 1b). Varying the PFO to protein and/or lipid ratio, the time or the temperature of solubilization in PFO did not lead to the appearance of higher-order Orai oligomers (Supplementary Fig. 4).

When Stim and Orai are co-expressed in *Drosophila* S2 cells, a greatly amplified CRAC current is recorded after store depletion[4]. To analyse Orai stoichiometry when the CRAC channel is functional, we used S2 cells transfected with Stim and Orai and performed chemical cross-linking experiments on total cell lysates of resting cells and cells that had been treated with thapsigargin to deplete the Ca^{2+} stores (Fig. 1c). When Orai was transfected alone, no obvious change in the cross-linking pattern was observed with or without thapsigargin treatment. In contrast, store depletion caused a decrease in the cross-linked dimer intensity of a V5-His-tagged Stim protein expressed alone, and a decrease in the intensity of both Orai and Stim dimers in co-transfected cells. Because a constant amount of protein was loaded into each lane, most of the Stim protein, and the Orai protein when co-expressed with Stim, were probably present as

[1]Department of Physiology and Biophysics, [2]Department of Neurobiology and Behavior, and [3]Center for Immunology, University of California Irvine, California 92697-4561, USA.

very high molecular mass cross-linked aggregates that did not enter the gel.

Because aggregation of Stim and Orai into macromolecular complexes precluded the determination of Orai stoichiometry in the active state, we sought a way to activate the CRAC channel without inducing higher order Orai cluster formation and, in addition, to test whether aggregation of Orai into punctae is a requirement for CRAC channel function. The C-terminal portion of STIM1, expressed as a cytosolic protein, activates CRAC current constitutively in Jurkat T cells[20] and in HEK cells co-transfected with Orai1 (refs 13 and 21). Expression of *Drosophila* C-terminal Stim bearing a V5-His tag (C-Stim) induced constitutive Ca^{2+} influx in S2 cells (Supplementary Fig. 5). We compared currents in S2 cells co-transfected with GFP–Orai and with either full-length Stim or C-Stim. When co-transfected with Stim, GFP–Orai produced an amplified CRAC current that developed on passive store depletion (Fig. 2a), which is similar in time course (half-time of 83 ± 24 s, $n = 11$ cells) but approximately tenfold larger in amplitude than native CRAC currents[22] and is consistent with results described previously for Stim plus untagged Orai[4]. In contrast, cells transfected with C-Stim plus GFP–Orai showed a robust pre-activated CRAC-like current immediately on breaking-in to initiate whole-cell recording (Fig. 2b and Supplementary Fig. 6a). The pre-activated current was identical to native or amplified CRAC currents in its inwardly rectifying *I–V* shape, its selectivity for Ca^{2+} and its sensitivity to block by 5 nM Gd^{3+} (Fig. 2a, b and Supplementary Fig. 6b, c). However, the pre-activated current induced by C-Stim plus GFP–Orai declined to a half-maximal value in 34 ± 9 s ($n = 14$ cells), probably owing to pipette dialysis resulting in dilution and unbinding of cytosolic C-Stim, which, unlike full length Stim, is not constrained to the ER membrane while interacting with Orai. Indeed, when small pipettes with higher series resistance

and correspondingly slower diffusional access were used, the decline of the pre-activated current proceeded more slowly (Supplementary Fig. 6d, e).

We next examined the localization of GFP–Orai in relation to co-expressed C-Stim or Stim (Fig. 2c and Supplementary Fig. 7). Immunostaining of cytosolic C-Stim showed it was present in a ring near the surface membrane and was co-localized with GFP–Orai, suggestive of a constitutive coupling. Notably, no punctae of C-Stim and co-expressed GFP–Orai were observed, even though CRAC channels were constitutively active. In contrast to this homogeneous distribution of C-Stim and co-expressed Orai, co-localized full-length Stim and Orai showed distinct punctae as expected after Ca^{2+} store depletion. The C-terminal portion of Stim includes the coiled–coil motif involved in dimerization but lacks the amino-terminal sterile alpha motif domain responsible for higher order Stim aggregation after store depletion. We have previously shown by co-immunoprecipitation that store depletion induces a dynamic coupling between Stim and Orai[12]. Here, we further demonstrate that *Drosophila* C-Stim physically interacts with Orai independently of the ER Ca^{2+} store content (Supplementary Fig. 8). Moreover, western blots after maleimide 1,6-bismaleimidohexane (BMH) cross-linking of intact cells demonstrated that C-Stim induces a shift towards higher order Orai homomultimers, including a clear tetramer population that was not observed when Orai was expressed alone (Fig. 2d). Collectively, these experiments show that C-Stim expressed as a cytoplasmic protein associates with and constitutively activates Orai subunits in the plasma membrane without forming punctae, and they provide evidence that the oligomeric state of Orai is dependent on C-Stim binding.

To determine Orai stoichiometry at the single-molecule level in the native membrane environment of intact cells, we used a recently
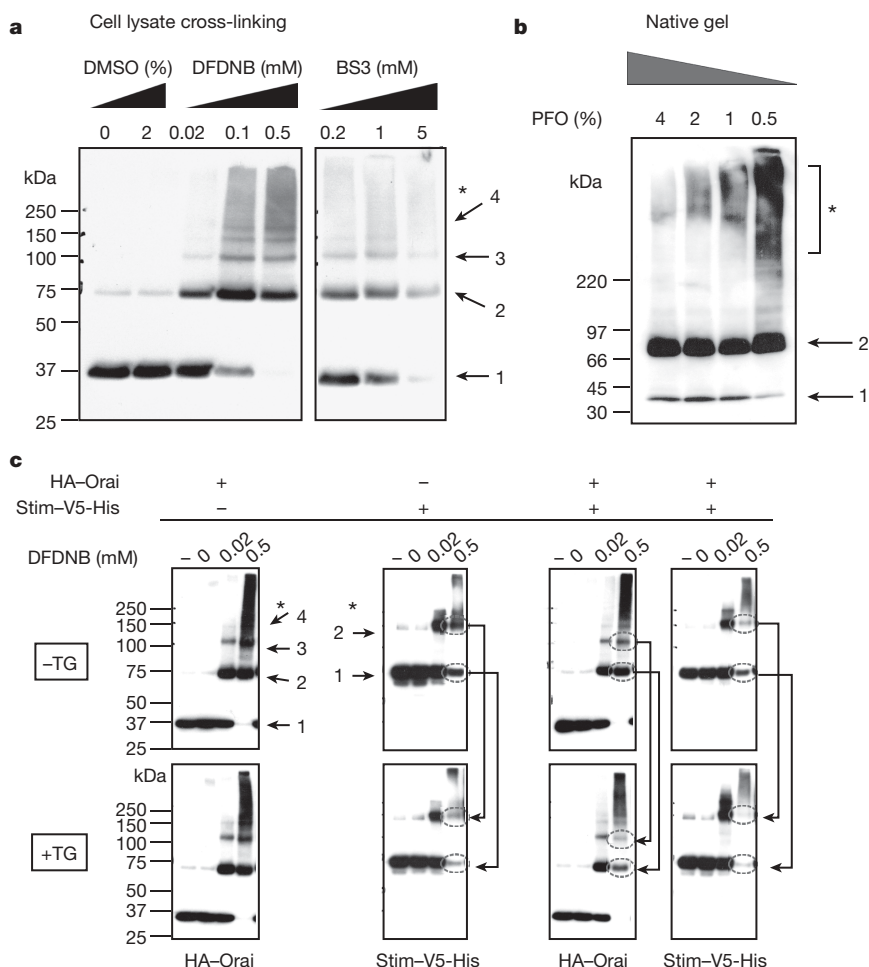
Figure 1 | Orai is mainly present as homodimers in resting S2 cells. Each panel is representative of at least 3 independent experiments. Numbers represent the assigned state of oligomerization: 1, monomer; 2, dimer; 3, trimer; 4, tetramer. Asterisks denote high-order aggregates.
a, Determination of Orai oligomeric structure using chemical cross-linking. DFDNB (1,5-difluoro-2,4-dinitrobenzene, membrane permeant) and BS3 (bis(sulfosuccinimidyl)suberate, membrane impermeant) were incubated with HA–Orai-transfected S2 cell lysates and the sizes of the cross-linked products were analysed by SDS–PAGE on 4–12% gradient gels. Orai oligomers, from dimer to tetramer, were observed with the dimer always being the predominant population. Similar results were obtained in intact cells using DSP (dithiobis(succinimidyl propionate), membrane permeant, data not shown), DFDNB (see Supplementary Figs 2 and 3) and the cysteine-reactive cross-linker BMH (membrane permeable). **b**, Confirmation of Orai dimerization using PFO–PAGE. HA–Orai-transfected S2 cell lysates were incubated with sample buffer containing different PFO concentrations for 30 min at room temperature before electrophoresis. **c**, DFDNB cross-linking of total cell lysates of S2 cells transfected with HA–Orai, with Stim–V5-His or co-transfected. Labels indicate untreated cells (−), dimethylsulphoxide (DMSO) vehicle control (0) and concentrations of DFDNB. Arrows show the reduction in Stim and Orai low-order oligomers upon Ca^{2+} store depletion by thapsigargin (1.5 μM for 15 min).

developed method[23] of total internal reflection microscopy (TIRFM) to image bleaching steps of individual GFP-tagged Orai. *Xenopus* oocytes were injected with complementary RNA (cRNA) for GFP–Orai with or without coincident injection of cRNA for Stim or C-Stim. When Stim was co-expressed, depletion of the Ca^{2+} store with thapsigargin resulted in clustering of GFP–Orai in punctae (Fig. 3a), making single molecules difficult to resolve. However, consistent with the results described earlier in *Drosophila* S2 cells, oocytes expressing both GFP–Orai and C-Stim showed a large Ca^{2+} influx as assessed by Ca^{2+} fluorimetry, together with the activation of an

endogenous Ca^{2+}-dependent Cl^- current, whereas neither Orai nor C-Stim alone were effective (Fig. 3b). Moreover, TIRFM imaging of GFP–Orai-expressing oocytes showed numerous diffraction-limited fluorescent spots (Fig. 3c) that were absent in the non-injected oocytes and increased in density with the time of expression. Continuous exposure to laser excitation resulted in stepwise decrements of fluorescence at these spots (Supplementary Videos 1 and 2) corresponding to bleaching of individual GFP molecules[23]. Different spots showed varying numbers of bleaching steps, ranging from one to a maximum of four (Fig. 4a, b). Notably, estimates of the mean
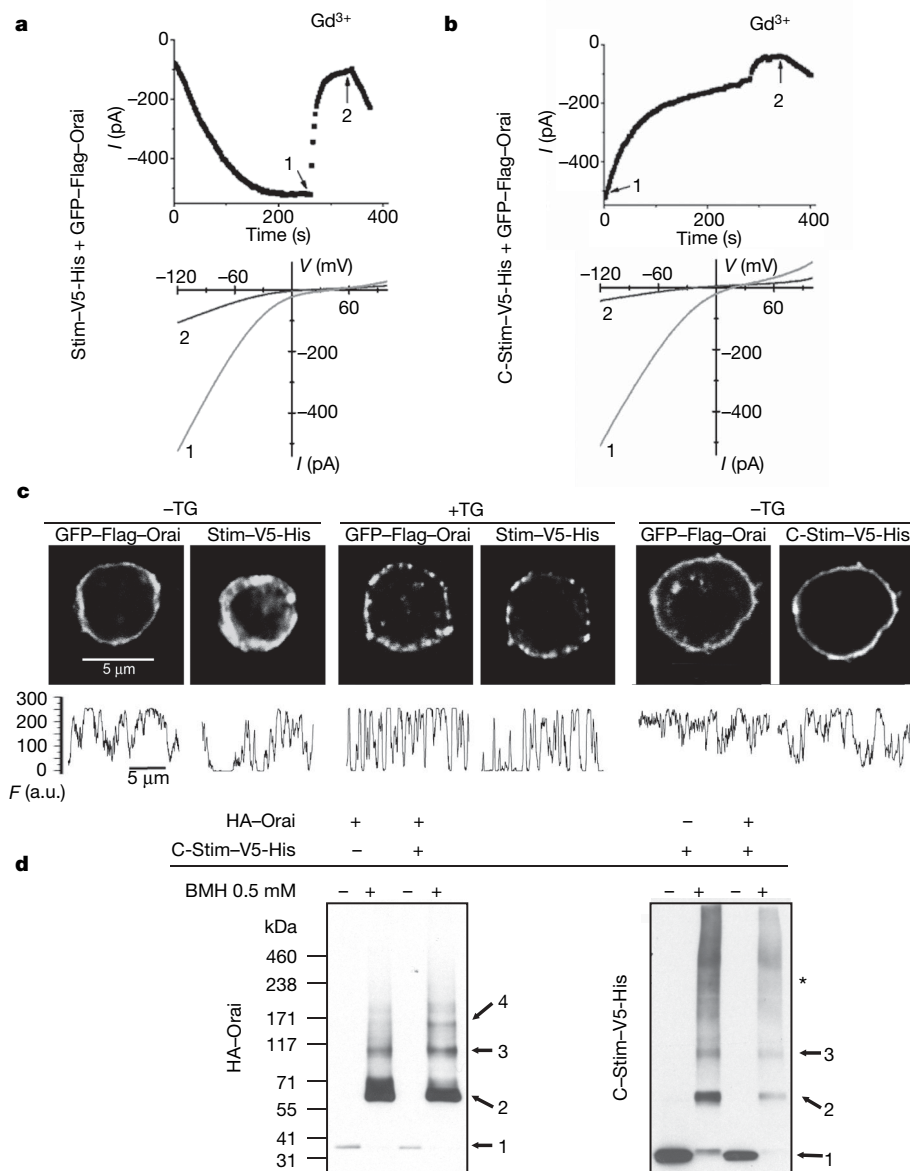

Figure 2 | The C terminus of Stim (C-Stim) constitutively activates Orai without forming punctae. a, b, Time course (top graphs) and *I–V* curves (bottom graphs) comparing the CRAC current in representative cells co-transfected with Stim + GFP–Orai (**a**) or with C-Stim + GFP–Orai (**b**). Application of 5 nM Gd^{3+} reversibly blocked most of the current in cells transfected either with Stim + GFP–Orai or with C-Stim + GFP–Orai. **c,** Subcellular localization of GFP–Orai together with Stim–V5-His or C-Stim–V5-His in resting (Stim–V5-His and C-Stim–V5-His) or store-depleted (2 μM thapsigargin for 15 min, Stim–V5-His only) co-transfected S2 cells. Puncta formation was only observed after store-depletion in cells expressing Stim–V5-His and GFP–Orai. The graphs underneath each picture show the fluorescence intensity profiles (*F*) for Orai and Stim obtained from the same regions of interest, tracing the perimeter of each cell clockwise from the top. a.u., arbitrary units. To quantify the extent to which

GFP–Orai was inhomogeneously distributed, we calculated the ratio of fluorescence variance to mean fluorescence from profiles such as those illustrated. Mean ratio values for GFP–Orai ($\pm$ s.e.m.; $n = 6$ cells for each condition) were: Stim–V5-His $-$ thapsigargin (TG), 29.03 ± 2.9; Stim–V5-His $+$ thapsigargin, 62.9 ± 4.5 ($P = 0.00002$); C-Stim–V5-His $-$ thapsigargin, 22.4 ± 3.7 (not significantly different from Stim–V5-His). **d,** BMH cross-linking in intact S2 cells transfected with HA–Orai, with C-Stim–V5-His or co-transfected with both. Numbers represent the inferred state of oligomerization: 1, monomer; 2, dimer; 3, trimer; 4, tetramer. The asterisk denotes higher order aggregates. The cross-linking pattern of Orai showed a clear tetrameric population when co-expressed with C-Stim but not in the absence of C-Stim. The cross-linking profile of C-Stim was not affected by the presence of Orai, and appeared mostly as dimers and trimers.

number of GFP molecules per spot made in this way differed markedly depending on the expression of C-Stim. In the oocytes expressing GFP–Orai alone, most of the spots (~70%) showed two steps to complete bleaching (Fig. 4a, c)—consistent with biochemical

observations in S2 cells—whereas after co-expression with C-Stim, most spots (~62%) showed four-step bleaching (Fig. 4b, c). The small proportions of spots that showed one- or three-step bleaching may reflect instances of near-simultaneous stochastic bleaching steps that could not be separately resolved, or expression of non-fluorescent GFP molecules[23]. The optical resolution of the microscope (approximately 250 nm) is inadequate to determine whether a spot showing four bleaching steps is truly a tetramer or, for example, two
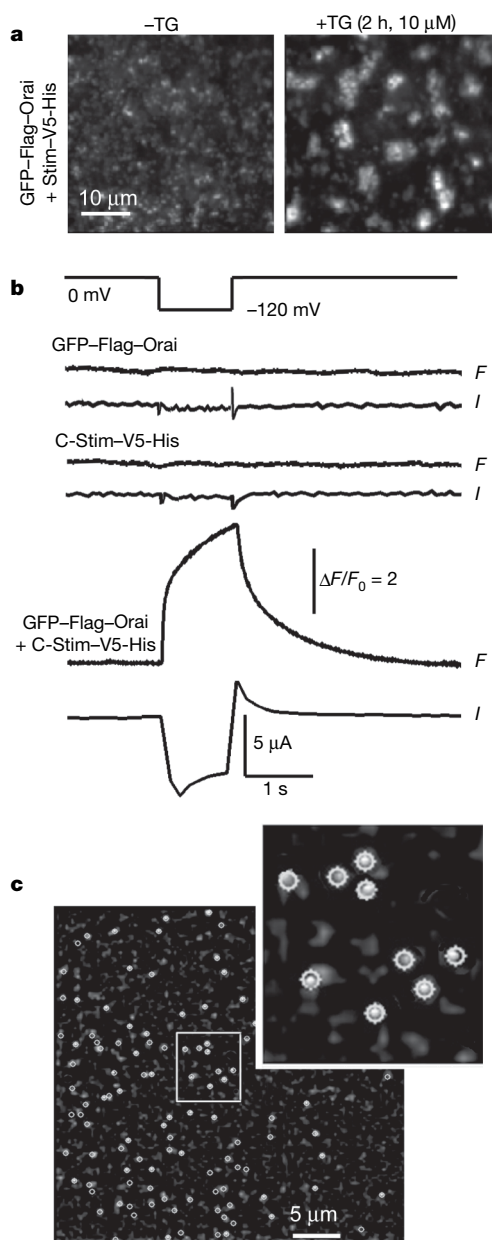

Figure 3 | Single-molecule photobleaching of GFP–Orai in intact oocytes. **a**, Store depletion induces the formation of Orai punctae in *Xenopus* oocytes. Images were obtained by TIRFM of oocytes expressing GFP–Orai together with Stim–V5-His and show $40 \times 40\,\mu m$ regions in the animal hemisphere before (left) and 2 h after (right) bath application of $10\,\mu M$ thapsigargin (TG) in zero-calcium Ringer's solution. **b**, Oocytes expressing GFP–Orai together with C-Stim showed strong Ca^{2+} influx, whereas this was absent with expression of GFP–Orai or C-Stim alone. The pairs of traces show cytosolic $[Ca^{2+}]$ as reported by normalized fluorescence pseudo-ratio changes of Fluo-4 (*F*) and whole-cell voltage-clamp measurements of Ca^{2+}-activated Cl^- current (*I*) in response to a hyperpolarizing step from 0 mV to -120 mV (top trace). **c**, Representative TIRFM image, acquired before photobleaching, showing fluorescent spots (circled) sparsely distributed in the membrane of an oocyte expressing GFP–Orai together with C-Stim. The inset shows a magnified view of a small region (white box), with circular regions of interest used to measure bleaching steps overlaid on the image.
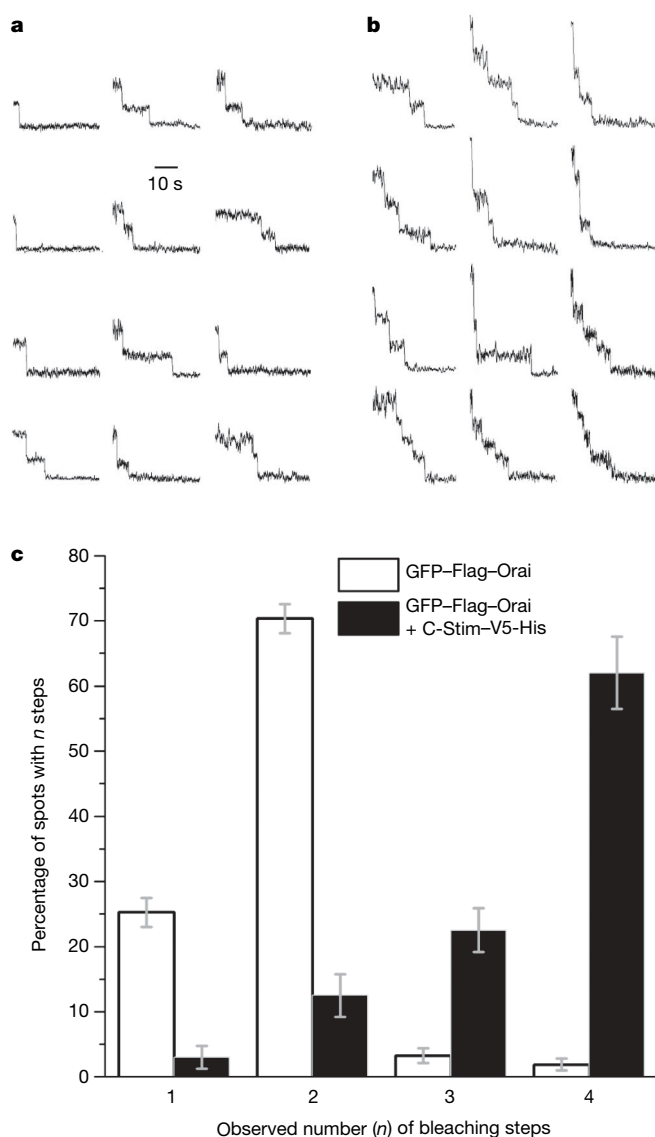

Figure 4 | GFP–Orai forms dimers in the basal state and mainly tetramers when co-expressed with C-Stim. **a**, **b**, Representative examples of single-molecule bleaching records obtained from oocytes expressing GFP–Orai alone (**a**) or GFP–Orai together with C-Stim (**b**). **c**, Histogram shows the percentages of spots that showed one, two, three and four bleaching steps in oocytes expressing GFP–Orai alone (open bars) and GFP–Orai plus C-Stim (filled bars). Errors bars indicate ± 1 s.e.m. Data for GFP–Orai were obtained from 400 spots, 11 imaging records and 6 oocytes; data for GFP–Orai + C-Stim were obtained from 278 spots, 5 imaging records and 3 oocytes. Comparison of bleaching step distributions with and without C-Stim yielded a Chi-square value of 590, $P < 0.001$. This cannot be attributed to an increased likelihood of two GFP–Orai dimers happening to occur indistinguishably close to each other owing to an increased expression level or to C-Stim-induced clustering, because fluorescence spots in both conditions showed similar random distributions and densities (respectively, 37 ± 6 and 41 ± 4 spots in a $40 \times 40\,\mu m^2$ region), and we did not observe spots with more than four bleaching steps as might be expected for a macromolecular clustering.

119

distinct dimers linked by C-Stim. We favour the former interpretation on the basis of the evidence[15] that the expression of an Orai1 tandem-tetramer construct forms functional CRAC channels, and that CRAC is inhibited when one subunit in the tetramer is replaced by a dominant-negative Orai. Thus, Orai is present in the membrane predominantly as dimers under basal conditions, and activation by C-Stim induces association to form tetramers.

Taken together, our results show that Orai adopts different quaternary structures depending on its activation state. In resting cells, Orai is present in the plasma membrane as a dimer, forming stable structural units, but when CRAC is activated, Orai is found predominantly as a tetramer. This result reconciles biochemical evidence pointing to a stable Orai dimer[14] (Fig. 1) with electrophysiological evidence from tandem constructs indicating a tetrameric channel[15]. Moreover, we show that the coiled–coil C-terminal domain of Stim is sufficient to trigger dimerization of Orai dimers to form the functional tetrameric channel and to activate CRAC influx. However, at present we cannot distinguish whether this dimer to tetramer transition is sufficient to activate Orai channel activity, or if a Stim-induced conformational change in each subunit of Orai is further required for channel activity. The channel assembly and activation mechanism identified here is mechanistically unique in its requirement for an activator protein (Stim) to assemble and open the Orai tetrameric channel in the plasma membrane.

Note added in proof: A recent study[24] reported a tetrameric Orai1 stoichiometry of the CRAC channel by Fluorescence imaging methods.

METHODS SUMMARY

Drosophila S2 cells (Invitrogen) and HEK293 cells (American Type Culture Collection, ATCC) were propagated and transfected (see complementary DNAs described in Methods) as described previously[2,12]. Chemical cross-linking was performed as described[17] with minor modifications (see Methods). Cross-linking experiments were also performed on living S2 cells directly incubated with different concentrations of cross-linkers. Protein complexes were fractionated by PFO–PAGE as described previously[17,19] and in Methods. Co-immunoprecipitations were performed in S2 cells as described[12] or on cells solubilized in PBS, 1% NP-40 and 5 mM EDTA for Stim–Orai interaction analysis. Equal amounts of protein were immunoprecipitated with the antibodies specified in the figure legends. After extensive washing, eluted samples were analysed by western blotting. Constructs and tags are described in Methods.

Transfected S2 cells were selected for whole-cell recording by fluorescence of GFP–Flag–Orai. Handling of C-Stim-transfected cells, solution recipes, voltage stimuli and data acquisition protocols are included in Methods.

Single-molecule bleaching experiments were performed on defolliculated *Xenopus* oocytes that had been injected 12–24 h previously with cRNAs for GFP–Orai alone or together with C-Stim–V5-His. Calcium influx was assayed by applying voltage-clamped hyperpolarizing pulses at the same time as monitoring the Ca^{2+}-activated Cl^- current and fluorescence of intracellularly loaded Fluo-4. Individual GFP–Orai multimers were visualized by TIRFM and image sequences were analysed by placing small regions of interest around fluorescence spots to manually count the bleaching steps during continuous exposure to 488 nm laser light.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Liou, J. *et al.* STIM is a Ca^{2+} sensor essential for Ca^{2+}-store-depletion-triggered Ca^{2+} influx. *Curr. Biol.* **15**, 1235–1241 (2005).
2. Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca^{2+} channel function. *J. Cell Biol.* **169**, 435–445 (2005).
3. Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).
4. Zhang, S. L. *et al.* Genome-wide RNAi screen of Ca^{2+} influx identifies genes that regulate Ca^{2+} release-activated Ca^{2+} channel activity. *Proc. Natl Acad. Sci. USA* **103**, 9357–9362 (2006).
5. Vig, M. *et al.* CRACM1 is a plasma membrane protein essential for store-operated Ca^{2+} entry. *Science* **312**, 1220–1223 (2006).
6. Cahalan, M. D. *et al.* Molecular basis of the CRAC channel. *Cell Calcium* **42**, 133–144 (2007).
7. Lewis, R. S. The molecular choreography of a store-operated calcium channel. *Nature* **446**, 284–287 (2007).
8. Zhang, S. L. *et al.* STIM1 is a Ca^{2+} sensor that activates CRAC channels and migrates from the Ca^{2+} store to the plasma membrane. *Nature* **437**, 902–905 (2005).
9. Wu, M. M., Buchanan, J., Luik, R. M. & Lewis, R. S. Ca^{2+} store depletion causes STIM1 to accumulate in ER regions closely associated with the plasma membrane. *J. Cell Biol.* **174**, 803–813 (2006).
10. Prakriya, M. *et al.* Orai1 is an essential pore subunit of the CRAC channel. *Nature* **443**, 230–233 (2006).
11. Vig, M. *et al.* CRACM1 multimers form the ion-selective pore of the CRAC channel. *Curr. Biol.* **16**, 2073–2079 (2006).
12. Yeromin, A. V. *et al.* Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai. *Nature* **443**, 226–229 (2006).
13. Muik, M. *et al.* Dynamic coupling of the putative coiled-coil domain of ORAI1 with STIM1 mediates ORAI1 channel activation. *J. Biol. Chem.* **283**, 8014–8022 (2008).
14. Gwack, Y. *et al.* Biochemical and functional characterization of Orai proteins. *J. Biol. Chem.* **282**, 16232–16243 (2007).
15. Mignen, O., Thompson, J. L. & Shuttleworth, T. J. Orai1 subunit stoichiometry of the mammalian CRAC channel pore. *J. Physiol. (Lond.)* **586**, 419–425 (2008).
16. Hoenderop, J. G. *et al.* Homo- and heterotetrameric architecture of the epithelial Ca^{2+} channels TRPV5 and TRPV6. *EMBO J.* **22**, 776–785 (2003).
17. Kedei, N. *et al.* Analysis of the native quaternary structure of vanilloid receptor 1. *J. Biol. Chem.* **276**, 28613–28619 (2001).
18. Raab-Graham, K. F. & Vandenberg, C. A. Tetrameric subunit structure of the native brain inwardly rectifying potassium channel K_{ir} 2.2. *J. Biol. Chem.* **273**, 19699–19707 (1998).
19. Ramjeesingh, M., Huan, L. J., Garami, E. & Bear, C. E. Novel method for evaluation of the oligomeric structure of membrane proteins. *Biochem. J.* **342**, 119–123 (1999).
20. Huang, G. N. *et al.* STIM1 carboxyl-terminus activates native SOC, I_{crac} and TRPC1 channels. *Nature Cell Biol.* **8**, 1003–1010 (2006).
21. Zhang, S. L. *et al.* Store-dependent and -independent modes regulating CRAC channel activity of human Orai1 and Orai3. *J. Biol. Chem.* **283**, 17662–17671 (2008).
22. Yeromin, A. V., Roos, J., Stauderman, K. A. & Cahalan, M. D. A store-operated calcium channel in *Drosophila* S2 cells. *J. Gen. Physiol.* **123**, 167–182 (2004).
23. Ulbrich, M. H. & Isacoff, E. Y. Subunit counting in membrane-bound proteins. *Nature Methods* **4**, 319–321 (2007).
24. Ji, W. *et al.* Functional stoichiometry of the unitary calcium-release-activated calcium channel. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0806499105 (29 August 2008).

Author Contributions A.P. designed and performed cDNA and cRNA constructs, cell biology and all biochemical experiments. A.D. performed all experiments in oocytes and data analysis of single-molecule photobleaching. A.V.Y. was responsible for patch-clamp experiments and analysis. S.L.Z. made the C-Stim construct, performed Ca^{2+} imaging experiments and collaborated on co-immunoprecipitation experiments. O.S. provided general technical assistance. I.P. supervised single-molecule experiments on oocytes. M.D.C. provided advice and overall direction, and supervised project planning and execution. A.P., I.P. and M.D.C. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.D.C. (mcahalan@uci.edu).

METHODS

Cell culture and transfection. *Drosophila* S2 cells (Invitrogen) were propagated and transfected as described previously[12]. Cells were used 16 h after transfection for immunocytochemistry and 36–96 h after transfection for patch-clamp, single-cell $[Ca^{2+}]_i$ imaging and biochemistry. HEK293 cells (ATCC) were maintained and propagated as recommended by the ATCC. HEK293 cells were transfected using Lipofectamine 2000 (Invitrogen) reagents and used after 48 h for biochemistry.

Molecular cloning and *in vitro* transcription. HA–Orai, Flag–Orai, Stim–V5-His and GFP constructs for *Drosophila* expression were described previously[12]. The GFP–Flag–Orai fusion protein (GFP–Orai) was made by introducing by PCR a 5′ in-frame EcoR1 site just after the starting methionine of the Flag–Orai coding sequence and ligating the fragment into the EcoR1–BamH1 sites of a monomeric variant of the enhanced GFP (pEGFP-C2, Clontech) in which the A206K mutation[25] was introduced by site-directed mutagenesis (Quickchange site-directed mutagenesis kit, Stratagene). The resulting GFP–Flag–Orai coding sequence was subcloned between the EcoR1 and the Nhe1 sites of the pAc5.1/V5-His B *Drosophila* expression vector (Invitrogen). The C-terminal fragment of Stim (C-Stim, amino acids 315–570) was generated by introducing by PCR a Not1 site followed by a methionine upstream of amino acid 315, and subcloning the resulting fragment between the Not1 and the Xho1 site of the pAc5.1/V5-His A vector (Invitrogen) as described previously for full-length Stim pAc5.1/D-STIM-V5-His (ref. 8). For HEK expression, *Drosophila* HA–Orai in pAc5.1/V5-His B was subcloned between the EcoR1 and the Xho1 sites of the pcDNA3.1/Zeo(+) mammalian expression vector (Invitrogen) and the rat Δ3N-P2X$_2$ receptor in pcDNA3, which was a gift from F.-A. Rassendren[26].

Xenopus expression constructs were obtained by subcloning the Not1–BamH1 GFP–Flag–Orai fragment between the corresponding sites of the pXLII vector[27] (gift from J. E. Hall), or by subcloning the Not1–Pme1 Stim–V5-His or C-Stim–V5-His fragments between the Not1–EcoR5 sites of PXLII. This vector contains the 5′- and 3′-untranslated regions of the *Xenopus laevis* β-globin gene with an internal multiple cloning site and a T3 promoter. Capped cRNAs were transcribed *in vitro* using T3 RNA polymerase (mMESSAGE mMACHINE kit, Ambion).

All mutants and constructs were verified by DNA sequencing on both strands and by analytical endonuclease restriction enzyme digestion; function was tested by whole cell patch-clamp recording and/or Ca^{2+} imaging.

Co-immunoprecipitation and western-blotting. Co-immunoprecipitation was performed in S2 cells as described[12] with some modifications. After treatment, 5–10×10^6 cells were lysed in 300 μl of either RIPA lysis buffer (Upstate) for Orai oligomerization experiments, or in PBS, 1% NP-40 and 5 mM EDTA for Stim–Orai co-immunoprecipitations, both supplemented with $1\times$ complete EDTA-free protease inhibitor mixture (Roche) and passed five times through a 26G needle. After 30 min of solubilization at 4 °C under agitation, lysates were centrifuged (16,000g, 10 min, 4 °C) and the supernatant was collected. Equal amounts of protein (250–500 μg) were diluted at 0.5 μg μl^{-1} in PBS and mixed with either anti-HA-probe monoclonal-antibody-conjugated agarose beads (5 μl beads per 100 μg total protein, Santa Cruz), anti-Flag M2 monoclonal-antibody-conjugated agarose beads (5 μl beads per 100 μg total protein, Sigma), anti-V5 monoclonal antibodies (1.25 μg per 100 μg total protein, Invitrogen), or anti-HA-probe monoclonal antibodies (1 μg per 100 μg total protein, Santa Cruz) overnight at 4 °C on a rotating wheel. For the non-conjugated antibodies, 25 μl of UltraLink protein A/G beads (Pierce) were subsequently added and rocked for 2 h at 4 °C. Beads were washed five times (5 min at 4 °C) with 1 ml of RIPA lysis buffer containing 10% glycerol or PBS and 0.05% NP-40, and proteins were eluted by boiling in $2\times$ concentrated LDS sample buffer (Invitrogen) supplemented with 100 mM dithiothreitol (DTT). Samples were resolved by SDS–PAGE on a 4–12% NuPAGE gradient gel (Invitrogen) and analysed by standard western blotting techniques. Cells transfected individually with each plasmid were used as controls, and immunodepletion of all samples was checked by SDS–PAGE of the flow through protein.

Immunoblots were incubated with the primary antibodies indicated, including: mouse anti-HA peroxidase-coupled antibody (Roche; 1:500) in PBS plus 0.5% casein blocking solution (Bio-Rad), for 1 h at room temperature; mouse anti-HA probe antibody (SantaCruz; 1:500) in PBS plus 0.5% casein, overnight at 4 °C; mouse anti-FlagM2 peroxidase-coupled antibody (Sigma; 1:2,500) in PBS plus 0.05% Tween-20, for 1 h at room temperature; mouse anti-α-tubulin DM1A antibody (Sigma; 1:2,000) in PBS plus 0.1% casein, for 2 h at room temperature; and mouse anti-V5 peroxidase-coupled antibody (Invitrogen; 1:5,000) in PBS plus 0.5% casein, for 1 h at room temperature. Proteins were detected by developing with the ECL+ detection kit (GE Healthcare).

Chemical cross-linking. The following homobifunctional reagents (Pierce) were used for protein cross-linking studies: the lysine-reactive *N*-hydroxysuccimide esters BS3 (water-soluble, non-membrane-permeable, spacer arm length 11.4 Å) and DSP (water-insoluble, membrane permeable, spacer arm length 12 Å), the lysine-reactive aryl halide DFDNB (water-insoluble, membrane permeable, spacer arm length 3 Å) and the cysteine-reactive BMH (water-insoluble, membrane permeable, spacer arm length 11.4 Å). All reagents were dissolved in PBS or DMSO immediately before use.

Chemical cross-linking was performed as described[17] with some modifications. The cell cultures were collected and rinsed twice with ice-cold PBS and then sonicated in ice-cold cross-linking buffer (PBS pH 8 for lysine-reactive cross-linkers or PBS pH 7.4 and 2.5 mM EDTA for cysteine-reactive reagent) containing protease inhibitors (complete-mini EDTA-free, Roche). The protein content was measured by a micromethod using the Bio-Rad protein assay. Fifteen micrograms of proteins in an equal volume of cross-linking buffer were incubated with different concentrations of cross-linkers indicated in the figure legend or with the same volume of vehicle. Incubation was performed for 10 min at 37 °C and stopped by the addition of 20 mM Tris, pH 7.5, for the lysine-reactive cross-linkers. Alternatively, for the cysteine-reactive BMH, the treatment was performed for 20 min at room temperature and ended by the addition of 25 mM DTT. The quenched samples were subsequently mixed with $4\times$ concentrated Nu-PAGE sample buffer containing 50 mM DTT (Invitrogen), incubated at 70 °C for 10 min, and then subjected to electrophoresis. The samples were separated on 4–12% gradient Nu-PAGE Bis-Tris gels or 3–8% gradient Nu-PAGE Tris-acetate gels and analysed by western blotting.

Cross-linking experiments were also performed on living S2 cells. Cells (5×10^6) were resuspended in 1 ml of cross-linking buffer and directly incubated with different concentrations of cross-linkers as described previously. After quenching, cell lysates were prepared and analysed by western blotting.

PFO–PAGE. Proteins complexes were fractionated by PFO–PAGE as described[17,19]. Total cell lysates were prepared as for chemical cross-linking experiments. In some experiments, cell were lysed in the presence of 1% of the mild NP-40 detergent for 20 min at 4 °C under agitation and centrifuged at 16,000g for 10 min to remove cellular debris. The lysates (30–40 μg) at 2 μg μl^{-1} were mixed with doubly concentrated PFO sample buffer (100 mM Tris-base, 2–8% NaPFO (Oakwood Products Inc.), 20% glycerol and 0.005% bromophenol blue, pH 8.0) plus 25 mM DTT. After 25 min of incubation at room temperature, the samples were vortex mixed, centrifuged for 5 min at 10,000g and then subjected to electrophoresis on 4–12% precast gradient Novex Tris-Glycine gels (Invitrogen) with a running buffer containing 25 mM Tris, pH 8.5, 192 mM glycine and 0.5% NaPFO, adjusted with sodium hydroxide and electroblotted as described[19]. As molecular mass standards, the high-molecular-mass rainbow marker kit (GE Healthcare), cross-linked albumin and phosphorylase *b* (Sigma) were resuspended in PFO sample buffer, separated on the same gels, electroblotted and stained with amido black (Sigma). These methods were validated by showing that under the same experimental conditions and cellular context, the rat P2X$_2$ channel assembled as a trimer-hexamer (Supplementary Fig. 3b), in accordance with its previously described stoichiometry[28].

Immunocytochemistry. After washing in calcium-free Ringer, transfected S2 cells on poly-L-lysine-coated glass coverslips were treated for 10 min at room temperature with 1.5 μM thapsigargin in calcium-free Ringer or left untreated in normal Ringer, fixed for 15 min at room temperature in 4% paraformaldehyde and 4% sucrose in PBS. Cells were washed in PBS containing 1% normal goat serum plus 50 mM glycine, and were then permeabilized in PBS containing 1% normal goat serum plus 0.05% Triton X-100. Blocking of nonspecific binding sites was performed by incubating cells with PBS containing 10% normal goat serum for 20 min and the primary antibody was added for 2 h at room temperature (anti-V5 mouse monoclonal antibody, Invitrogen; 1:200). After an extensive wash, cells were incubated for 25 min at 37 °C with Alexa-Fluor-conjugated secondary antibody (Alexa594 goat anti-mouse antibody, Invitrogen; 1:2,000). Stained cells were viewed by confocal microscopy (Zeiss LSM510 META). Background staining was determined by incubating non-transfected cells with both primary and secondary antibodies (data not shown).

Single-cell $[Ca^{2+}]_i$ imaging. Ratiometric Fura-2 $[Ca^{2+}]_i$ imaging was performed in S2 cells as described[2,8]. Cells transfected with C-Stim were recognized by co-transfected GFP, with appropriate filters used to avoid contamination of Fura-2 fluorescence by bleed-through of GFP fluorescence. Data were analysed using METAFLUOR software (Molecular Devices) and ORIGINPRO 7.5 software (OriginLab) and are expressed as means ± s.e.m.

Whole-cell recording. Patch-clamp experiments were performed in S2 cells at room temperature in the standard whole-cell recording configuration, as described[12]. To decrease the Ca^{2+} influx caused by a preactivated CRAC current that might damage cells before recording or immunostaining, S2 cells co-transfected with C-Stim plus GFP–Orai, and control cells transfected with Stim plus GFP–Orai, were maintained for 16–96 h in complete calcium-free Schneider's insect medium (Sigma) and plated on poly-L-lysine-coated coverslips. For

whole-cell recording, cells were maintained in nominally Ca^{2+}-free external solution. After seal formation and just before break-in to achieve whole-cell recording, the standard 2 mM Ca^{2+} solution was applied locally to the cell. Pipette resistances were normally ~2 MΩ, but pipettes ranging from 8 to 12 MΩ were used to evaluate diffusional access as a mechanism for run-down of constitutive CRAC currents induced by C-Stim expression. The recipes of external and internal solutions are indicated in Supplementary Table 1. Only cells with high input resistance (>2 GΩ) were selected for recording. Membrane potentials were corrected for a liquid junction potential of 10 mV between the pipette and the bath solution. The series resistance was not compensated. The membrane potential was held at 0 mV, and 220-ms voltage ramps from −120 to 100 mV alternating with 220-ms pulses to −120 mV were delivered every 2 s. To calculate current densities, peak current amplitudes were divided by the membrane capacitance for each cell.

Oocyte preparation and cRNA injection. Plasmids containing cDNA clones coding for the *Drosophila* GFP–Flag–Orai and C-Stim–V5-His subunits were linearized and transcribed *in vitro*, and cRNAs were mixed to a final concentration of 0.001–0.01 μg μl^{-1} and injected (30 nl) into defolliculated stage VI oocytes obtained from *Xenopus laevis*[29]. After injection, oocytes were maintained for 12–24 h in Barth's solution (1.8 mM Ca^{2+}), and were then prepared for TIRFM imaging by manual removal of the vitelline membrane after shrinking in hypertonic 'stripping' solution (composition in mM: potassium aspartate, 200; KCl, 20; MgCl$_2$, 1; EGTA, 10; HEPES, 10; pH 7.2, cooled to 4 °C).

Oocyte electrophysiology and Ca^{2+} measurement. Ca^{2+} influx was assayed using a two-electrode voltage clamp to apply hyperpolarizing steps to oocytes bathed in Ringer's solution with [Ca^{2+}] raised to 6 mM, and simultaneously measuring Ca^{2+}-activated Cl$^-$ current[30] and monitoring cytosolic free [Ca^{2+}] by means of the fluorescence of Fluo-4 dextran loaded to a final intracellular concentration of ~ 40 μM[31].

Single-molecule GFP photobleaching and analysis. TIRFM of single GFP–Orai molecules expressed in *Xenopus* oocytes was accomplished using a home-built system[29] based around an Olympus IX70 microscope equipped with a ×60, NA 1.45 TIRF objective. Devitellinated oocytes were allowed to settle on a cover glass forming the base of the recording chamber and were bathed in calcium-free Ringer's solution (composition in mM: NaCl, 120; KCl, 2; MgCl$_2$, 5; EGTA, 1, HEPES, 5; at pH 7.4). GFP-tagged molecules lying within the ~100 nm evanescent field were excited by total internal reflection of a 488 nm laser beam incident through the microscope objective. Images (128 × 128 pixel; 1 pixel = 0.33 μm) were acquired at 10 frames s^{-1} by a Cascade 128+ electron multiplying CCD camera (Roper Scientific). The resulting image stacks (1,000 frames; 100 s) were processed in MetaMorph (Molecular Devices) by averaging every two consecutive frames, followed by subtraction of a heavily smoothed (7 × 7 pixel) copy of each frame so as to correct for bleaching of autofluorescence and other background signals. Traces such as those in Fig. 4a, b were then obtained around selected spots, excluding those that showed obvious movement or where spots were too close to be unambiguously separated. The number of bleaching steps in each trace was determined by visual inspection, with measurements restricted to those spots that showed complete bleaching and where fluorescence steps could be clearly resolved above the noise level.

25. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).

26. Newbolt, A. *et al.* Membrane topology of an ATP-gated ion channel (P2X receptor). *J. Biol. Chem.* **273**, 15177–15182 (1998).

27. Nielsen, P. A. *et al.* Molecular cloning, functional expression, and tissue distribution of a novel human gap junction-forming protein, connexin-31.9. Interaction with zona occludens protein-1. *J. Biol. Chem.* **277**, 38272–38283 (2002).

28. Aschrafi, A., Sadtler, S., Niculescu, C., Rettinger, J. & Schmalzing, G. Trimeric architecture of homomeric P2X$_2$ and heteromeric P2X$_{1+2}$ receptor subtypes. *J. Mol. Biol.* **342**, 333–343 (2004).

29. Demuro, A. & Parker, I. Optical single-channel recording: imaging Ca^{2+} flux through individual ion channels with high temporal and spatial resolution. *J. Biomed. Opt.* **10**, 11002 (2005).

30. Parker, I., Gundersen, C. B. & Miledi, R. A transient inward current elicited by hyperpolarization during serotonin activation in *Xenopus* oocytes. *Proc. R. Soc. Lond. B* **223**, 279–292 (1985).

31. Dargan, S. L., Demuro, A. & Parker, I. Imaging Ca^{2+} signals in *Xenopus* oocytes. *Methods Mol. Biol.* **322**, 103–119 (2006).

Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications

Lauren G. Holden[1]*, Courtney Prochnow[1]*, Y. Paul Chang[1]*, Ronda Bransteitter[1], Linda Chelico[1], Udayaditya Sen[1], Raymond C. Stevens[2], Myron F. Goodman[1] & Xiaojiang S. Chen[1]

The APOBEC family members are involved in diverse biological functions. APOBEC3G restricts the replication of human immunodeficiency virus (HIV), hepatitis B virus and retroelements by cytidine deamination on single-stranded DNA or by RNA binding[1–4]. Here we report the high-resolution crystal structure of the carboxy-terminal deaminase domain of APOBEC3G (APOBEC3G-CD2) purified from *Escherichia coli*. The APOBEC3G-CD2 structure has a five-stranded β-sheet core that is common to all known deaminase structures and closely resembles the structure of another APOBEC protein, APOBEC2 (ref. 5). A comparison of APOBEC3G-CD2 with other deaminase structures shows a structural conservation of the active-site loops that are directly involved in substrate binding. In the X-ray structure, these APOBEC3G active-site loops form a continuous 'substrate groove' around the active centre. The orientation of this putative substrate groove differs markedly (by 90 degrees) from the groove predicted by the NMR structure[6]. We have introduced mutations around the groove, and have identified residues involved in substrate specificity, single-stranded DNA binding and deaminase activity. These results provide a basis for understanding the underlying mechanisms of substrate specificity for the APOBEC family.

We have purified the human wild-type C-terminal cytidine deaminase domain of APOBEC3G (APOBEC3G-CD2, residues 197–380) expressed in *E. coli*. APOBEC3G-CD2 (with and without a glutathione *S*-transferase (GST) tag) is highly soluble, and deaminates cytidine to uridine on single-stranded DNA (ssDNA) with a specific activity of 5 fmol μg^{-1} min^{-1}, which is about 25-fold lower than that of the full-length APOBEC3G (GST–APOBEC3G; 126 fmol μg^{-1} min^{-1}) expressed in *E. coli* (Fig. 1a). We analysed the processive and polar properties of APOBEC3G-CD2 and full-length APOBEC3G (Fig. 1b). Similar to the insect-cell-derived full-length APOBEC3G[7,8], the full-length APOBEC3G expressed in *E. coli* processively deaminates cytidine in two 5′CCC3′ motifs located on a ssDNA substrate, during one binding event (Fig. 1b). The full-length APOBEC3G also exerts a 3′ to 5′ deamination bias by preferentially deaminating the cytidine in the CCC motif near the 5′ end of the ssDNA substrate (Fig. 1b). In contrast, the APOBEC3G-CD2 exhibits an approximate twofold decrease in processivity and virtually no 3′ to 5′ deamination bias (Fig. 1b). These results indicate that APOBEC3G-CD2 partially retains the catalytic properties of full-length APOBEC3G, but that the CD1 domain in the context of the full-length APOBEC3G is probably required for the strong processive property and the 3′ to 5′ deamination bias on ssDNA.

We crystallized the wild-type APOBEC3G-CD2 and solved the structure by using the multi-wavelength anomalous dispersion (MAD) phasing method with selenium-substituted methionine

(Se-Met) diffraction data. The 2.3 Å resolution X-ray structure shows a core β-sheet that is composed of five β-strands surrounded by six α-helices (Fig. 1c, d). Helices 2–4 (h2–h4) are packed alongside one face of the core β-sheet (Fig. 1c), whereas helix 1 (h1) and helix 5 (h5) are packed against the opposite face of the β-sheet (Fig. 1c, d). Helix 6 (h6) is located at the edge of the β-sheet core and is perpendicular to the β5 strand (Fig. 1c).

A recently reported NMR structure of an APOBEC3G-CD2 mutant (APOBEC3G-2K3A, Protein Data Bank (PDB) accession 2JYW)[6] resembles the X-ray structure of the wild-type APOBEC3G-CD2. However, the superposition of the two structures
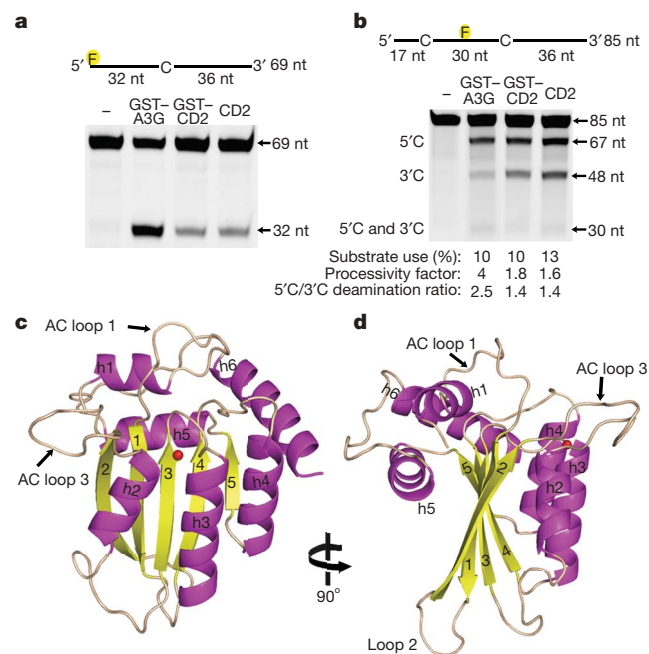

Figure 1 | The X-ray structure of enzymatically active APOBEC3G-CD2.
a, Analysis of the deamination activity for full-length GST–APOBEC3G (GST–A3G) and GST–APOBEC3G-CD2 (GST–CD2) and APOBEC3G-CD2 (CD2). The 32-nucleotide (nt) band indicates deamination activity. F represents the position of fluorescein-dT incorporated into the ssDNA. **b**, APOBEC3G processivity and the 3′ to 5′ deamination bias was characterized on ssDNA with two CCC motifs. Single deaminations of the 5′C and the 3′C appear as 67- and 48-nucleotide fragments, respectively; deamination of both the 5′C and the 3′C results in a 30-nucleotide fragment (see Methods) **c, d**, Two views of the APOBEC3G-CD2 domain rotated 90° showing the five-stranded β-sheet core surrounded by six helices. The zinc is represented as a red sphere.

[1]Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089, USA. [2]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
*These authors contributed equally to this work.

shows notable differences (Supplementary Fig. 1a). The β2 strand and the amino-terminal helix (h1) are absent from the NMR structure (Supplementary Fig. 1b, c, in grey). The β2 strand in the X-ray structure does not make crystal contact with neighbouring monomers. Thus, the formation of the β2 strand is unlikely to be the result of crystal contact. Furthermore, a similar β2 strand within a five-stranded β-sheet core is the common structural feature that is observed in all wild-type cytidine deaminase structures available so far (Supplementary Fig. 2a, b, d, e). Therefore, an intact full-length β2 strand and the five-stranded β-sheet core is probably the feature of wild-type APOBEC3G-CD2 and all other APOBEC proteins. The structural differences observed in the NMR structure could have resulted from the five mutations on the APOBEC3G protein used for NMR study (Supplementary Fig. 1b, c), or from the different methodology used for determining the structure, or from both.

A superposition of the core structures of APOBEC3G-CD2 and APOBEC2 monomers shows substantial overlap for all five β-strands and all six helices (Fig. 2a), suggesting that these core structures of APOBEC family members are highly conserved. Yet, the structural overlap shows notable differences in the active centre (AC) loops,
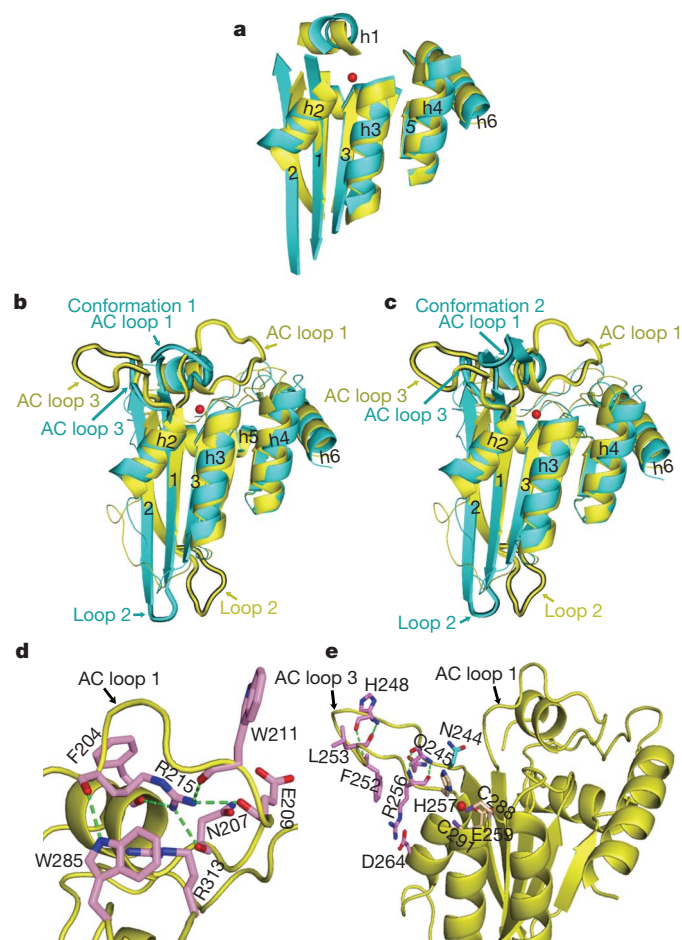
referred to as AC loops 1 and 3, which potentially mark the differences in substrate use and activity of the two proteins (Fig. 2b, c). The AC loop 1, which connects h1 with the β1 strand, is located further away from the active site in APOBEC3G than in APOBEC2 (Fig. 2b, c). The APOBEC3G AC loop 3 is longer than that of APOBEC2 and is positioned further away from the active site (Fig. 2b, c).

The structure shows elaborate bonding interactions that can stabilize the open conformation of APOBEC3G AC loops 1 and 3. For example, AC loop 1 forms an extensive bonding network through R215, which anchors this loop to other parts of the structure (Fig. 2d). R215 interactions include the direct contact with R313 and W285 located near the core structure (Fig. 2d). We demonstrate later that the R215E mutation in APOBEC3G abolishes deamination activity, which is consistent with a previous study[9]. Similarly, the APOBEC3G AC loop 3 is stabilized by multiple hydrogen bonds between the main-chain atoms of residues R256, F252, L253, H248 and Q245 within the loop (Fig. 2e). The loop residue R256 interacts with D264 on a core helix by a strong salt bridge, and R256 hydrophobically packs with the loop residue F252 by the long aliphatic chain (Fig. 2e). All of these interactions should help stabilize the conformation of AC loop 3. Shown later, an APOBEC3G R256E mutant, which probably disrupts the AC loop 3 conformation, greatly impairs deamination activity.

In the active site of APOBEC3G-CD2, a zinc atom is coordinated by the three residues H257, C288 and C291, and a water molecule
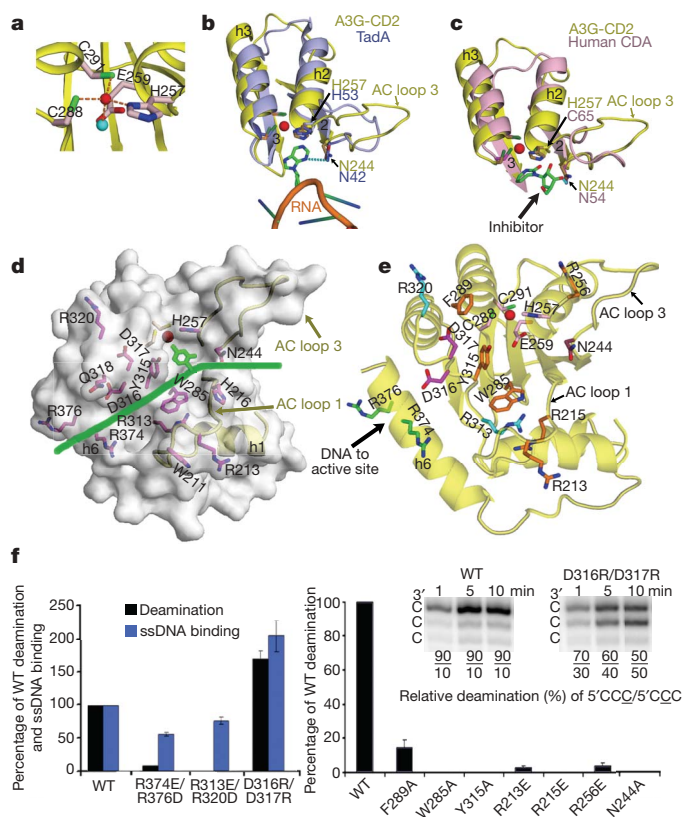

Figure 2 | Structural comparison of APOBEC3G-CD2 with APOBEC2.
a, Core structures of APOBEC3G-CD2 (yellow) and APOBEC2 (cyan) superimposed. The red sphere represents zinc. **b, c**, The superposition of APOBEC3G-CD2 and an APOBEC2 monomer, with the AC loop 1 collapsed over the active site (conformation 1, **b**) or forming an α-hairpin (conformation 2, **c**). **d**, In APOBEC3G-CD2, the AC loop 1 R215 residue forms hydrogen bonds (green dashes) with F204, W211, N207, E209 and W285 (pink). The R215 aliphatic chain hydrophobically packs with F204, R313 and W285. **e**, The APOBEC3G-CD2 AC loop 3 residues, R256, F252, L253, H248 and Q245 (pink), form main-chain hydrogen bonds (green dashes). The conserved N244 is shown in cyan. The active site residues are H257, C288 and C291 (wheat).

Figure 3 | Predicted substrate groove and deamination activity of APOBEC3G mutants. **a**, The active site residues of APOBEC3G-CD2. The water and zinc molecules are cyan and red spheres, respectively. **b**, Superposition of APOBEC3G-CD2 (A3G-CD2, yellow) and TadA (light blue, PDB accession 2B3J). **c**, Superposition of APOBEC3G-CD2 (yellow) and human CDA (pink, PDB accession 1MQo). **d**, Surface representation of APOBEC3G-CD2, showing a horizontal groove with residues (magenta) predicted to interact with ssDNA. ssDNA is represented by a green line. **f**, Mutational data of APOBEC3G purified from Sf9 (left) or from *E. coli* (right) are shown. The right inset shows the relative deamination of the 3′C (5′CCC) or the middle C (5′CCC) on a ssDNA substrate by Sf9 purified proteins. Error bars represent the s.d.

(Fig. 3a). The closely positioned water molecule can be activated to become a Zn-hydroxide for nucleophilic attack in the deamination reaction[10]. Two residues (N244 and H257) on the APOBEC3G AC loop 3 show a structural conservation with many distantly related Zn-deaminases, specifically TadA and human CDA[10,11] (Fig. 3b, c). The two equivalent TadA residues (N42 and H53) on a TadA loop (similar to the AC loop 3) directly contact the target base of the RNA substrate (Fig. 3b). These residues overlap well with the APOBEC3G residues N244 and H257 on the AC loop 3 in the superposition of the two structures (Fig. 3b)[11]. Similarly, two equivalent residues (N54 and C65) on a human CDA loop contact the substrate/inhibitor[10], and also overlap with N244 and H257 on the AC loop 3 of APOBEC3G (Fig. 3c). This structural conservation suggests that the APOBEC3G-CD2 residues, N244 and H257, are also involved in substrate contact. In an *in vitro* assay, the APOBEC3G N244A mutant had no detectable deamination activity (Fig. 3f). The structural conservation of the position of these residues suggests that the open conformation of the APOBEC3G AC loop 3 is in a position ready to bind nucleic acid.

A surface representation of the APOBEC3G-CD2 X-ray structure shows that the AC loops 1 and 3 and the regions near the active site form a deep, spacious groove that runs horizontally across the active centre pocket (Fig. 3d). This groove is not present in the APOBEC3G-2K3A NMR structure because of the structural differences[6] (Supplementary Fig. 3d–f). The structural features in this groove strongly suggest a role for binding ssDNA substrates. The groove starts between the AC loops 1 and 3 on the right side of the displayed structure (Fig. 3d), leads into a deep pocket where the Zn atom is located and slightly exposed, and continues towards the left side over helix 6. The target base must be positioned into the active site so that the attachment of the Zn hydroxyl group can occur on the cytidine base during the deamination reaction (Fig. 3c). The ssDNA lying across this horizontal groove can present a cytidine base so that it is directed towards the active site Zn in the correct orientation and angle to permit deamination, as shown in the case of TadA and human CDA (Fig. 3a–c)[10,11].

Within this horizontal groove are a group of charged residues (R213, R215, N244, R256, R313, D316, D317, R320, R374 and R376) and hydrophobic residues (W285, Y315 and F289; Fig. 3e). In our mutagenesis study, we show that all of these residues are important for the deamination activity on ssDNA (Fig. 3f). However, they affect the deamination activity in different ways. The R374 and R376 residues are located on one end of the groove and are positioned to interact with a negatively charged ssDNA phosphate backbone. The ssDNA binding of the R374E/R376D double mutant is impaired by 46% in comparison to that of the wild-type APOBEC3G, and the deamination activity is even more disrupted (Fig. 3f). On the edge of the groove, the AC loop 1 R213 residue can make contact with ssDNA. Consistent with a previous report[6], the R213E mutant has only weak deamination activity (Fig. 3f).

Three of the charged residues (R256, R215 and R313) are involved in elaborate bonding networks for the AC loops (Figs 2d, e and 3e), and should be important for maintaining the groove conformation. The mutants R215E, R256E and R313E/R320D show only minimal or no deamination activity (Fig. 3f). The primary functional role of these residues may be to maintain the conformation of the substrate groove rather than to directly contact ssDNA. Mutation of the R313 residue can disrupt its interaction with W285, which is located on the floor of the groove near the active site Zn (Fig. 3e). Y315 next to W285 is also on the floor of the groove. Both residues could stack with bases of ssDNA and position the DNA into the active site (Fig. 3d, e). Mutants W285A and Y315A show no detectable deamination activity (Fig. 3f), consistent with a previous report[9]. Another hydrophobic residue on the edge of the groove is F289, and the F289A mutant has greatly reduced deaminase activity (Fig. 3d–f).

Notably, next to Y315 and W285 are two negatively charged residues (D316 and D317) on the floor of the groove (Fig. 3d, e). The mutant D316R/D317R has higher deamination activity (1.6-fold), as well as higher ssDNA binding (twofold) compared to the wild-type APOBEC3G (Fig. 3f). These enhanced activities could be caused by the increased total positive charge in the groove. Furthermore, this mutant showed altered substrate specificity (Fig. 3f, inset). Unlike wild-type APOBEC3G that strongly favours deamination at the 3′C of a 5′CCC3′ hot-spot motif, the D316R/D317R mutant deaminates the middle C and the 3′C at about the same rate (Fig. 3f, inset). This result indicates that these negative residues, D316 and D317, are important for positioning the substrate so that the 3′C is most likely to be deaminated by wild-type APOBEC3G.

The mutagenesis study supports our model of the horizontal groove, and verifies that the residues located within and around the groove are important for deamination activity, ssDNA binding and substrate orientation. These results provide a basis to pursue further studies of APOBEC3G and other important APOBEC proteins (including activation-induced cytidine deaminase, AID), which will facilitate our understanding of how they act within our innate and adaptive immune responses to restrict HIV and other infectious pathogens.

METHODS SUMMARY

APOBEC3G-CD2 was expressed and purified as a recombinant GST-fusion protein in *E. coli*. Purified GST-fusion protein was digested by PreScission protease. Further purification of the APOBEC3G-CD2 protein was completed using Superdex-75 gel filtration chromatography in 50 mM HEPES, pH 7.0, 250 mM NaCl and 1 mM dithiothreitol. Native and Se-Met-labelled proteins were concentrated to 25 mg ml^{-1}. Crystals were grown at 18 °C by hanging-drop vapour diffusion from a reservoir solution of 100 mM MES, pH 6.5, 40% PEG 200. In an assay for deamination activity, APOBEC3G (0.024–10 μM) was allowed to react with 500 nM fluorescein-dT-incorporated ssDNA for 10 or 15 min and subsequently treated with uracil-DNA glycosylase and resolved on 16% urea–PAGE for analysis as described previously[7]. Specific activity, measured as fmoles of substrate deaminated per μg of enzyme per minute, was calculated from the per cent deamination of a ssDNA substrate over a range of enzyme concentrations. To analyse processivity and directionality, substrate use (%) was less than 15% to maintain single-hit kinetics. The 'processivity factor' is defined as the ratio of the observed fraction of double deaminations (occurring at both 5′C and 3′C on the same molecule) to the predicted fraction of independent double deaminations[7]. A processivity factor of greater than one indicates that most of the double deaminations are caused by the same APOBEC3G molecule acting processively on both C targets. The deamination bias is measured by the ratio of 5′C/3′C deaminations[7]. For the experiments measuring processivity and directionality, the ssDNA substrate sequence are found in the Methods. Crystallography statistics are found in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Peled, J. U. *et al.* The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* **26,** 481–511 (2008).
2. Goodman, M. F., Scharff, M. D. & Romesberg, F. E. AID-initiated purposeful mutations in immunoglobulin genes. *Adv. Immunol.* **94,** 127–155 (2007).
3. Conticello, S. G., Langlois, M. A., Yang, Z. & Neuberger, M. S. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.* **94,** 37–73 (2007).
4. Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26,** 317–353 (2008).
5. Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445,** 447–451 (2007).
6. Chen, K. M. *et al.* Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452,** 116–119 (2008).
7. Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. *Nature Struct. Mol. Biol.* **13,** 392–399 (2006).
8. Chelico, L., Sacho, E. J., Erie, D. A. & Goodman, M. F. A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *J. Biol. Chem.* **283,** 13780–13791 (2008).
9. Chen, K. M. *et al.* Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. *FEBS Lett.* **581,** 4761–4766 (2007).

10. Chung, S. J., Fromme, J. C. & Verdine, G. L. Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.* **48,** 658–660 (2005).

11. Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nature Struct. Mol. Biol.* **13,** 153–159 (2006).

METHODS

Structure determination and refinement. Selenium-substituted methionine protein crystals were used for collecting Se-MAD data using the ALS synchrotron beam source. Data were processed with HKL3000 (ref. 12). A total of three selenium and one zinc sites were located by the SHELXD[13] program using MAD data between 50 and 3.0 Å resolution range. The SHARP program was used to calculate the experimental and model-combined phases using the MAD data in the resolution range of 50 to 2.3 Å as well as for density modification. The model was built with O using the high quality electron density map obtained, and was refined with CNS to 2.3 Å resolution with excellent statistics. The final refinement statistics and geometry as defined by Procheck were in good agreement and are summarized in Supplementary Table 1. Structure figures were designed using PyMOL[14].

Construction of APOBEC3G mutants. Mutant APOBEC3G proteins (D316R/D317R, R313E/R320D and R374E/R376D) were constructed by site-directed mutagenesis using the pAcG2T-APOBEC3G vector as the template. The following primers and their complementary strands were used: 5′-CTTCACTG-CCCGCATCTATAGAAGACAAGGAAGATGTCAGGAG-3′ (D316R/D317R), 5′-CTGTGCATCTTCACTGCCGAGATCTATGATGATCAAGGAGATTGTC-AGGAGGGGCTGCGC-3′ (R313E/R320D), and 5′-GAGCACAGCCAAGAC-CTGAGTGGGGAGCTGGACGCCATTCTCCAGAATCAGG-3′ (R374E/R376D). The entire coding region of the APOBEC3G mutant constructs was verified by DNA sequencing. The mutant plasmids were then co-transfected, according to the manufacturer's protocol, with linearized baculovirus DNA (BD Biosciences) to generate recombinant mutant APOBEC3G baculovirus. Wild-type and mutant APOBEC3G expression in Sf9 insect cells and purification was carried out as described previously[8]. Mutant *E. coli* GST–APOBEC3G proteins (R213E, R215E, K249E, R256E, W285A, F289A, Y315A and N244A) were constructed by site-directed mutagenesis using the pGEX-6P1-GST-APOBEC3G vector as the template. The following primers and their complementary strands were used: 5′-AATGAACCTTGGGTTGAAGGTCGTCACGAGACTTAC-3′ (R213E), 5′-GA-ACCTTGGGTTCGTGGTGAACACGAGACTTACCTG-3′ (R215E), 5′-TGTAAC-CAGGCCCCGCACGAGCACGGTTTTCTGGAA-3′ (K249E), 5′-GCACGGTT-TTCTGGAAGGTGAACACGCCGAACTGTG-3′ (R256E), 5′-GTTACCTGCTT-TACCTCTGCGTCCCCGTGCTTTTCC-3′ (W285A), 5′-ACCTCTTGGTCCC-CGTGCGCTTCCTGCGCACAAGAA-3′ (F289A), 5′-ATCTTCACTGCACGTA-TTGCCGACGACCAGGGCCGT-3′ (Y315A), 5′-CGTCGTGGTTTCCTGTGT-GCCCAGGCCCCGCACAAGCAC-3′ (N244A), 5′-CGTCGTGGTTTCCTGTC-TAGACAGGCCCCGCACAAGCAC-3′ (N244A). The entire coding region of the APOBEC3G mutant constructs was verified by DNA sequencing. Plasmids were expressed in XA90 *E. coli* cells and were lysed by French press. Further purification was carried out as described previously[8].

DNA binding. APOBEC3G DNA binding was monitored by changes in steady state fluorescence depolarization (rotational anisotropy). Reaction mixtures (70 μl), containing fluorescein-labelled DNA (50 nM) in buffer (50 mM HEPES, pH 7.3, 1 mM dithiothreitol and 5 mM $MgCl_2$) and varying concentration of 0 to 500 nM APOBEC3G, were incubated at 37 °C. The sequence of the ssDNA was TTAGATGAGTGTAA(fluorescein-dT)GTGATATATGTGTAT. Rotational anisotropy was measured as described previously[7]. The fraction of DNA bound to protein was determined as described previously[15].

Processivity and directionality substrates. The substrate used to determine processivity and directionality is 5′-AAAGAGAAAGTGATACCCAAAGAGT-AAAGT (fluorescein-dT)AGATAGAGAGTGATACCCAAAGAGTAAAGTTAG-TAAGATGTGTAAGTATGTTAA-3′. For specific activity measurements, the ssDNA substrate sequence was GG(fluorescein-dT)AGTTTAGTGGTTTGTAT-AGAATTAATACCCAAAGAAGTGTATGTAATTGTTATGATAAGATTGAAA.

12. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
13. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
14. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
15. Bertram, J. G., Bloom, L. B., O'Donnell, M. & Goodman, M. F. Increased dNTP binding affinity reveals a nonprocessive role for *Escherichia coli* β clamp with DNA polymerase IV. *J. Biol. Chem.* **279**, 33047–33050 (2004).

LETTERS

Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks

Daniel Zilberman[1], Devin Coleman-Derr[1], Tracy Ballinger[2,3] & Steven Henikoff[2,3]

Eukaryotic chromatin is separated into functional domains differentiated by post-translational histone modifications, histone variants and DNA methylation[1–6]. Methylation is associated with repression of transcriptional initiation in plants and animals, and is frequently found in transposable elements. Proper methylation patterns are crucial for eukaryotic development[4,5], and aberrant methylation-induced silencing of tumour suppressor genes is a common feature of human cancer[7]. In contrast to methylation, the histone variant H2A.Z is preferentially deposited by the Swr1 ATPase complex near 5′ ends of genes where it promotes transcriptional competence[8–20]. How DNA methylation and H2A.Z influence transcription remains largely unknown. Here we show that in the plant *Arabidopsis thaliana* regions of DNA methylation are quantitatively deficient in H2A.Z. Exclusion of H2A.Z is seen at sites of DNA methylation in the bodies of actively transcribed genes and in methylated transposons. Mutation of the MET1 DNA methyltransferase, which causes both losses and gains of DNA methylation[4,5], engenders opposite changes (gains and losses) in H2A.Z deposition, whereas mutation of the PIE1 subunit of the Swr1 complex that deposits H2A.Z[17] leads to genome-wide hypermethylation. Our findings indicate that DNA methylation can influence chromatin structure and effect gene silencing by excluding H2A.Z, and that H2A.Z protects genes from DNA methylation.

To investigate H2A.Z deposition in plant chromatin, we generated a high-resolution genome-wide map of H2A.Z in *Arabidopsis* by adapting the *in vivo* biotinylation system that we used to affinity purify *Drosophila* chromatin[21]. We tagged *Arabidopsis* H2A.Z with a peptide specifically recognized by the *Escherichia coli* biotin ligase BirA (biotin ligase recognition peptide, BLRP), and created transgenic plants coexpressing BLRP–H2A.Z with BirA. Cytological localization revealed that BLRP–H2A.Z has a diffuse nuclear distribution, but is excluded from heterochromatic chromocentres (Supplementary Fig. 1), the same pattern as that of endogenous H2A.Z[17]. After digestion with micrococcal nuclease to mostly mononucleosomes (Supplementary Fig. 1), we purified biotinylated chromatin from root tissue and co-hybridized the associated DNA with control DNA on tiling microarrays representing the entire *Arabidopsis* genome[22]. To ensure that our results were not influenced by potential tagging artefacts, we repeated the experiment with antibodies against endogenous H2A.Z[17]. We also mapped DNA methylation in roots (we have previously published a data set from aerial tissues[22]).

The maps generated by streptavidin pull-down and immunoprecipitation were virtually the same (Fig. 1 and Supplementary Fig. 2). The most notable feature was a strong, quantitative anticorrelation with DNA methylation (Pearson's $r = -0.81$; Supplementary Tables 1 and 2). Distinct peaks of H2A.Z around the 5′ ends of genes were also evident (Fig. 1b). To visualize better the H2A.Z distribution, we aligned all *Arabidopsis* annotated sequences, which include genes,

pseudogenes and transposable elements, at their 5′ ends, and stacked them from the top of chromosome 1 to the bottom of chromosome 5 (Fig. 2a and Supplementary Fig. 2). An obvious feature of this alignment is a vertical strip of high H2A.Z that roughly corresponds to the first nucleosome after the start of transcription. This pattern of H2A.Z deposition is consistent with those in yeast and humans[10–15], indicating that this is a general feature of eukaryotic genes. There were also five conspicuous horizontal stripes of low H2A.Z incorporation. These correspond to transposon-rich, heavily methylated heterochromatin surrounding the five *Arabidopsis* centromeres. This pattern of incorporation is precisely the opposite of that of DNA methylation (Fig. 2b and Supplementary Fig. 2).

Methylation is not distributed evenly within the genome. Transposons are heavily and uniformly methylated, some genes have short stretches of methylation, and most genes are unmethylated[22–26]. These three groups of sequences display a corresponding triphasic distribution of H2A.Z signal: low H2A.Z levels are found in transposons, intermediate levels in methylated genes, and high levels in unmethylated genes (Supplementary Fig. 3). One possibility is that the low levels of H2A.Z in transposons are caused by intrinsic sequence preferences, rather than DNA methylation. To test this, we examined the small number (49) of *Arabidopsis* transposons that are not methylated (Supplementary Table 3). Tellingly, all such transposons had high H2A.Z levels, indicating that low H2A.Z incorporation is not a feature of transposons *per se* (Figs 1c and 2c, d). Unmethylated transposons also lacked any discernible H2A.Z peaks, suggesting that these are unique features of endogenous genes. Unsupervised k-means clustering of annotated *Arabidopsis* sequences on the basis of H2A.Z patterns produced three groups that closely correspond to unmethylated genes, body-methylated genes and transposons (Fig. 2e, Supplementary Fig. 4 and Supplementary Table 4). Again, H2A.Z and DNA methylation levels showed a notable anticorrelation (Fig. 2f). DNA methylation and H2A.Z are thus mutually exclusive chromatin features, and our analyses show that this relationship is independent of sequence context, transcription or transcription potential (Supplementary Information and Supplementary Figs 5–11).

So far, our results indicate a strong anticorrelation between methylation and H2A.Z deposition, but we cannot distinguish which is causal. To address this issue, we took advantage of a line bearing a null mutation in the DNA methyltransferase *MET1*, *met1-6* (refs 4 and 27). Mutations in *MET1* cause major reductions in overall DNA methylation, and also significant hypermethylation mediated by other methyltransferases[26]. We reasoned that if DNA methylation influences H2A.Z deposition, changes in DNA methylation should be mirrored by changes in H2A.Z distribution. Notably, because *met1* causes both losses and gains of DNA methylation, we should see both gains and losses of H2A.Z. To test our hypothesis, we

[1]University of California, 211 Koshland Hall, Berkeley, California 94720, USA. [2]Fred Hutchinson Cancer Research Center, and [3]Howard Hughes Medical Institute, 1100 Fairview Avenue North, Seattle, Washington 98109, USA.
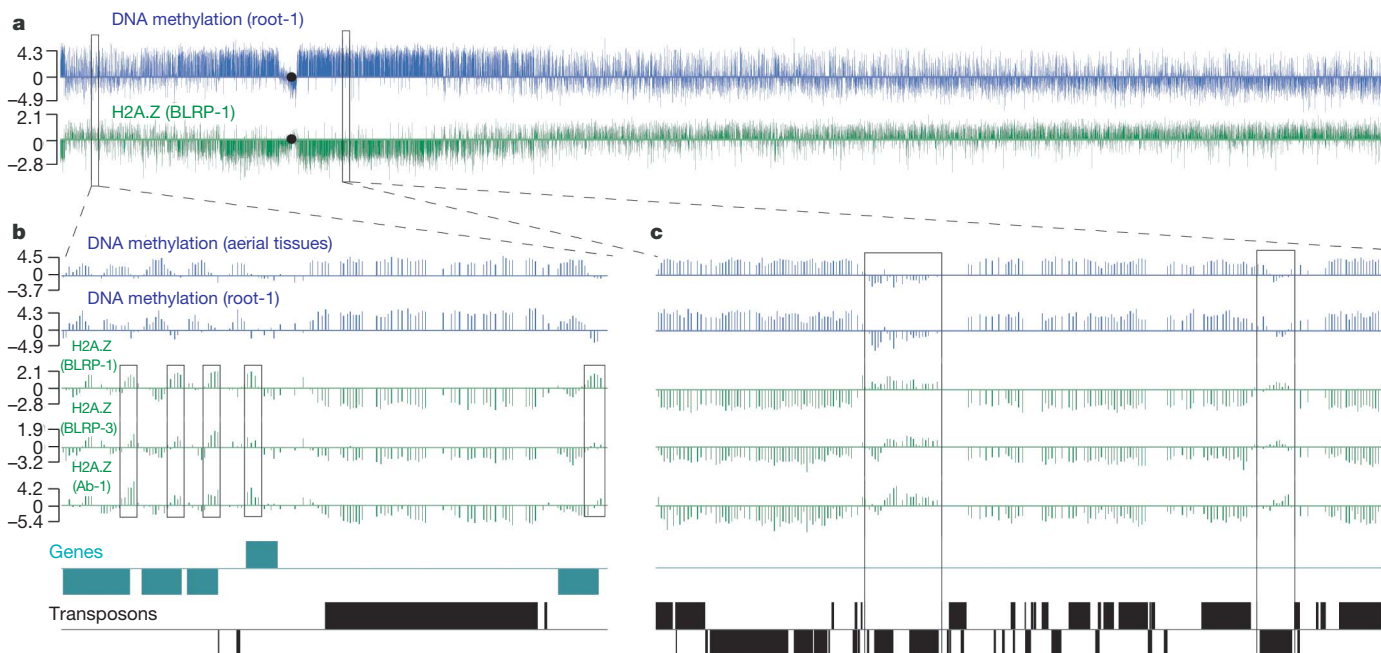
Figure 1 | High-resolution maps of *Arabidopsis* H2A.Z and DNA methylation. a, H2A.Z (green) and DNA methylation (blue) profiles of *Arabidopsis* chromosome 2. Each vertical bar represents the $\log_2$ signal ratio of the test sample signal divided by the input control signal. The black circles denote the position of the centromeric sequence gap. **b, c**, More detailed views of a euchromatic (positions 547,000–587,000, **b**) and a heterochromatic (4,407,000–4,463,000, **c**) genomic region. DNA

methylation from aerial tissues and roots is shown in blue; H2A.Z profiles obtained from two independent BLRP–H2A.Z transgenic lines and by means of immunoprecipitation of endogenous H2A.Z (Ab-1) are shown in green. Genes and transposons on the top and the bottom strands are shown above and below the line, respectively. 5′ peaks of H2A.Z in genes are emphasized by boxes in **b**. Unmethylated transposons with relatively high levels of H2A.Z are emphasized by boxes in **c**.

mapped H2A.Z, as well as DNA methylation and transcription, in *met1-6* plants.

Changes in DNA methylation indeed engendered changes in H2A.Z distribution (Fig. 3 and Supplementary Figs 12 and 13). To visualize these changes, we subtracted the wild-type H2A.Z data set from the *met1* H2A.Z data set, so that high values represent increased H2A.Z incorporation in *met1* (Supplementary Fig. 12). Examples of informative loci are shown in Fig. 3a–c. The *FWA* gene, which normally has 5′ methylation and lacks an H2A.Z peak, loses promoter methylation and gains 5′ H2A.Z in *met1* (Fig. 3a). The retrotransposon At5g13205

is heavily methylated in wild type, but loses methylation and gains H2A.Z in *met1* (Fig. 3b). Gene At1g22000, which encodes an F-box protein, is hypermethylated in *met1*, leading to loss of its 5′ H2A.Z peak (Fig. 3c).

To get a comprehensive view of H2A.Z dynamics in *met1-6*, we aligned and arranged all annotated *Arabidopsis* sequences as in Fig. 2a. The same conspicuous pericentric stripes were evident in this profile (Fig. 3d and Supplementary Fig. 13)—H2A.Z levels are increased in transposable elements, which lose most of their methylation and become reactivated in *met1* (refs 22 and 23). Unbiased sorting of
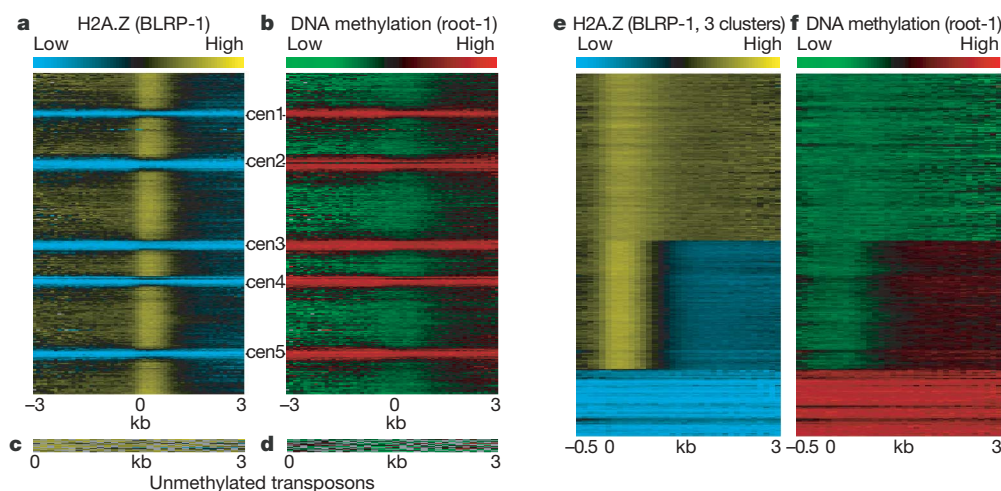

Figure 2 | H2A.Z and DNA methylation are mutually exclusive. a, b, All annotated sequences from the *Arabidopsis* Information Resource release 7 (TAIR7, 31,762 sequences) were aligned at the 5′ end and stacked from the top of chromosome 1 to the bottom of chromosome 5. BLRP–H2A.Z is displayed as a heat map in **a**; root DNA methylation is displayed in **b**. Centromeric gap positions are indicated (cen1 to cen5). Note the high degree of anticorrelation between H2A.Z and methylation.

c, d, Unmethylated transposable elements (listed in Supplementary Table 3). BLRP–H2A.Z is displayed as a heat map in **c**; root DNA methylation is displayed in **d**. **e**, All TAIR7 annotated sequences were *k*-means clustered (*k* = 3) on the basis of BLRP–H2A.Z patterns, and displayed as a heat map. For comparison, root DNA methylation of the same sequences is shown as a heat map in **f**.

126

the data produced three clusters that roughly encompass unmethylated genes, methylated genes and transposons, respectively (Fig. 3e, Supplementary Fig. 13 and Supplementary Table 4; sequences are categorized as in ref. 22). The changes in H2A.Z closely correspond to DNA methylation—sequences that gain H2A.Z in *met1* are methylated in wild type (Fig. 3f and Supplementary Fig. 13). Conversely, loci with decreased H2A.Z incorporation are unmethylated in wild type, but methylated in *met1-6* (Fig. 3g). Overall, changes in DNA methylation were mirrored by changes in H2A.Z in a manner that strongly argues that methylation inhibits H2A.Z incorporation.

Because some transposons and genes undergo transcriptional upregulation in *met1* plants[22], we had an opportunity to test whether H2A.Z incorporation is negatively influenced by methylation or positively influenced by transcription. Within genes, there is a robust correlation between DNA methylation in wild type and H2A.Z changes in *met1-6* (average Pearson's $r = 0.51$, Supplementary Table 2), but there is no correlation between transcriptional and H2A.Z changes (average Pearson's $r = 0.05$). *FWA*, which is strongly overexpressed in *met1*, has reduced levels of H2A.Z in the body of the gene, where it has no methylation in wild type (Fig. 3a). Similarly, of the handful of transposons that are not methylated in wild type, two

(At4g10690 and At5g35205) are nevertheless upregulated in *met1* (Supplementary Fig. 14). Both also have less H2A.Z in *met1* than in wild type, the opposite of other transposons.

Because only about half of all transposable elements are upregulated in *met1*, we could ask whether those elements preferentially gain H2A.Z, as would be expected if H2A.Z incorporation was associated with transcriptional activity. To ensure that the size of the data sets and methylation are not an issue, we compared 12,500 probes that represent activated transposons and silent transposons, respectively, and have identical methylation profiles. We found that both transposon classes are equally enriched in H2A.Z (Fig. 3h and Supplementary Fig. 15). Thus, changes in DNA methylation, rather than transcription, cause the redistribution of H2A.Z we observe in *met1*.

Our results show that DNA methylation excludes H2A.Z. An intriguing question is whether H2A.Z can also exclude methylation. Some of our data suggest that this is indeed the case. The most notable feature of H2A.Z incorporation, the 5′ genic peak, is independent of DNA methylation (Fig. 2e, f and Supplementary Figs 4 and 6), yet methylation is strongly excluded from this area[22,23]. Likewise, the higher H2A.Z levels in the bodies of less-transcribed genes (Supplementary
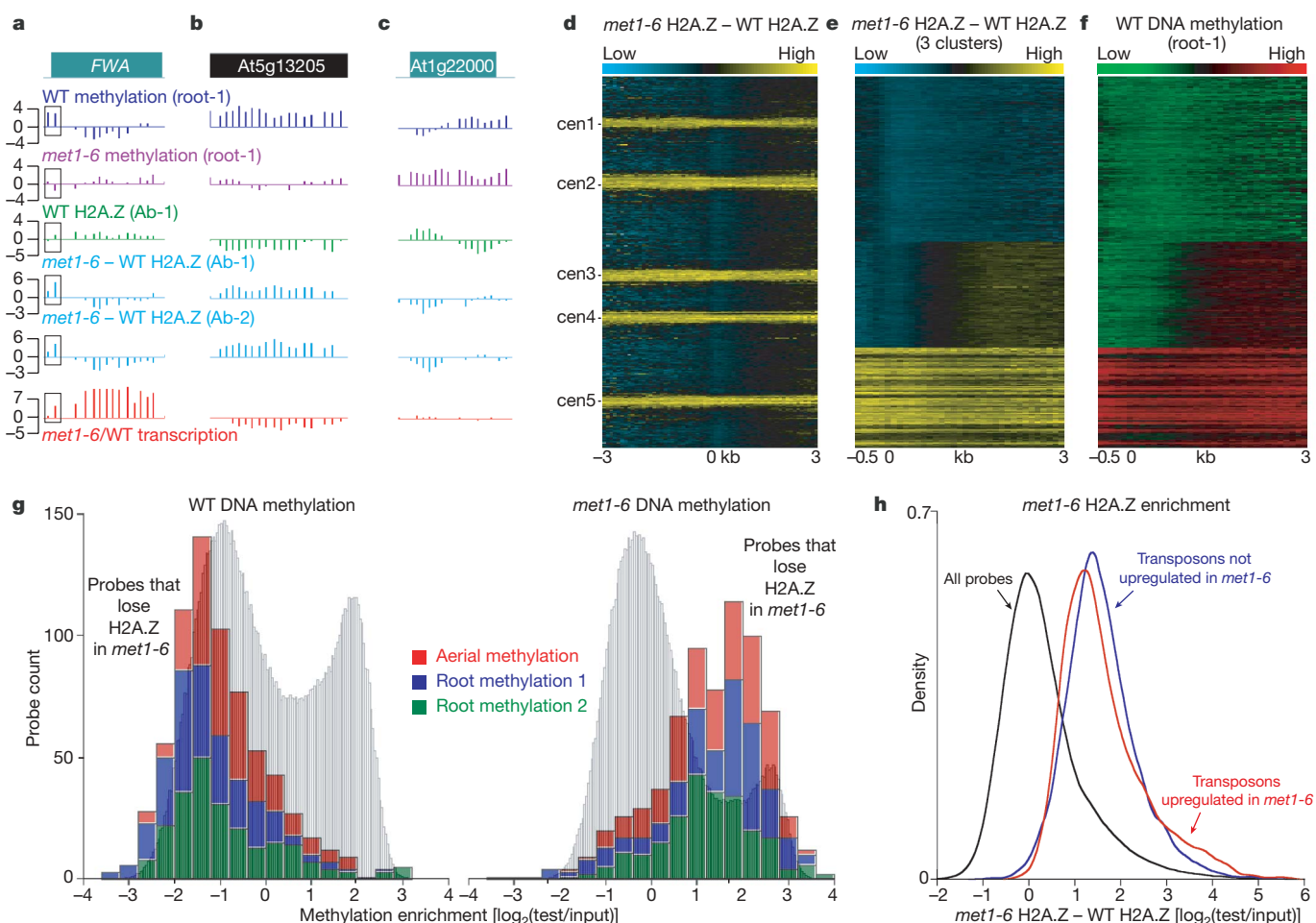

Figure 3 | H2A.Z incorporation changes in *met1-6* mutant plants. a–c, Wild type (WT) root DNA methylation (dark blue), *met1-6* root DNA methylation (purple), WT H2A.Z (antibody, green), WT H2A.Z profile subtracted from the *met1-6* H2A.Z profile (two sets of independent paired experiments, light blue), and *met1-6*/WT transcription (red) for *FWA* in **a,** copia-like transposable element At5g13205 that loses methylation and gains H2A.Z in *met1-6* in **b,** and F-box gene At1g22000 that is hypermethylated and loses H2A.Z in *met1-6* in **c.** The 5′ region of *FWA* methylated in WT is emphasized by boxes in **a.** d, All TAIR7 annotated sequences were aligned at the 5′ end and stacked from the top of chromosome 1 to the bottom of chromosome 5. The WT H2A.Z pattern

subtracted from the *met1-6* H2A.Z pattern is displayed as a heat map. The same data after *k*-means clustering (*k* = 3) are shown in **e.** For comparison, root DNA methylation of sequences arranged as in **e** is shown as a heat map in **f.** g, WT methylation levels (left) and *met1-6* methylation levels (right) for probes representing a significant decrease of H2A.Z in *met1-6* (Supplementary Fig. 12). The histogram is cumulative for three independent methylation data sets. Grey histograms in the background show the signal distribution for all probes. **h,** Kernel density plot, which has the effect of tracing the frequency distribution, of all probes in the data set displayed in **d** (black trace), transposable elements upregulated in *met1-6* (red trace), and transposable elements not upregulated in *met1-6* (blue trace).

Information and Supplementary Figs 6 and 7) might explain the puzzling observation that the chances of a gene becoming methylated increase with transcription (up to about the 70th percentile)[22,23].

To address this issue, we mapped DNA methylation in plants with a strong loss-of-function allele of *PIE1* (the conserved catalytic component of Swr1) that disrupts proper deposition of H2A.Z[17]. The overall methylation pattern in *pie1-5* plants remained similar to that in wild type (Supplementary Table 5), but there was a modest but consistent increase in DNA methylation (Supplementary Fig. 16). To visualize the methylation changes in *pie1*, we subtracted the methylation patterns of matched wild-type controls (F$_2$ siblings) from *pie1* and displayed the resulting data as a heat map (Fig. 4a and Supplementary Fig. 16). This analysis revealed genome-wide hypermethylation of gene bodies. Using the ChIPOTle algorithm[28], we identified 1,201 hypermethylated regions (corresponding to 1,172 genes) for further analysis (threshold $P < 10^{-7}$, Supplementary Table 6).

In plants, DNA methylation can occur at any cytosine[5]. Most methylation is found in symmetric CG sites, like it is in animals, and is mediated by MET1, but there is also a substantial amount of methylation in other sequence contexts catalysed by other methyltransferases

(hence the hypermethylation observed in *met1*)[25,26]. To determine how the *pie1* mutation affects DNA methylation in different contexts, we used bisulphite sequencing to analyse the methylation of individual cytosines in five loci scored as hypermethylated by ChIPOTle: At1g69850 (a nitrate transporter), At3g22340 (a COPIA-like retrotransposon), At4g03480 (an ankyrin-repeat-containing protein), At4g38190 (a cellulose synthase) and At5g37450 (a protein kinase). All five showed a modest but consistent gain of CG methylation (Fig. 4b, c), confirming the microarray analysis. There was very little non-CG methylation at any of the loci in either wild type or *pie1* (data not shown). Interestingly, all of the loci had some methylation in wild type, so the overall genomic hypermethylation we observed in *pie1* is likely to be primarily caused by increased methylation of normally lightly methylated loci rather than *de novo* methylation of previously unmethylated loci.

Given the wide-spread hypermethylation caused by the *pie1* mutation, we asked whether the hypermethylated loci are representative of the genome as a whole. As might be expected, *pie1* hypermethylated genes have high levels of H2A.Z in wild type (that is, those generally found in unmethylated genes; Fig. 4d). They are also generally enriched in genes transcribed at a low level, with greatest enrichment
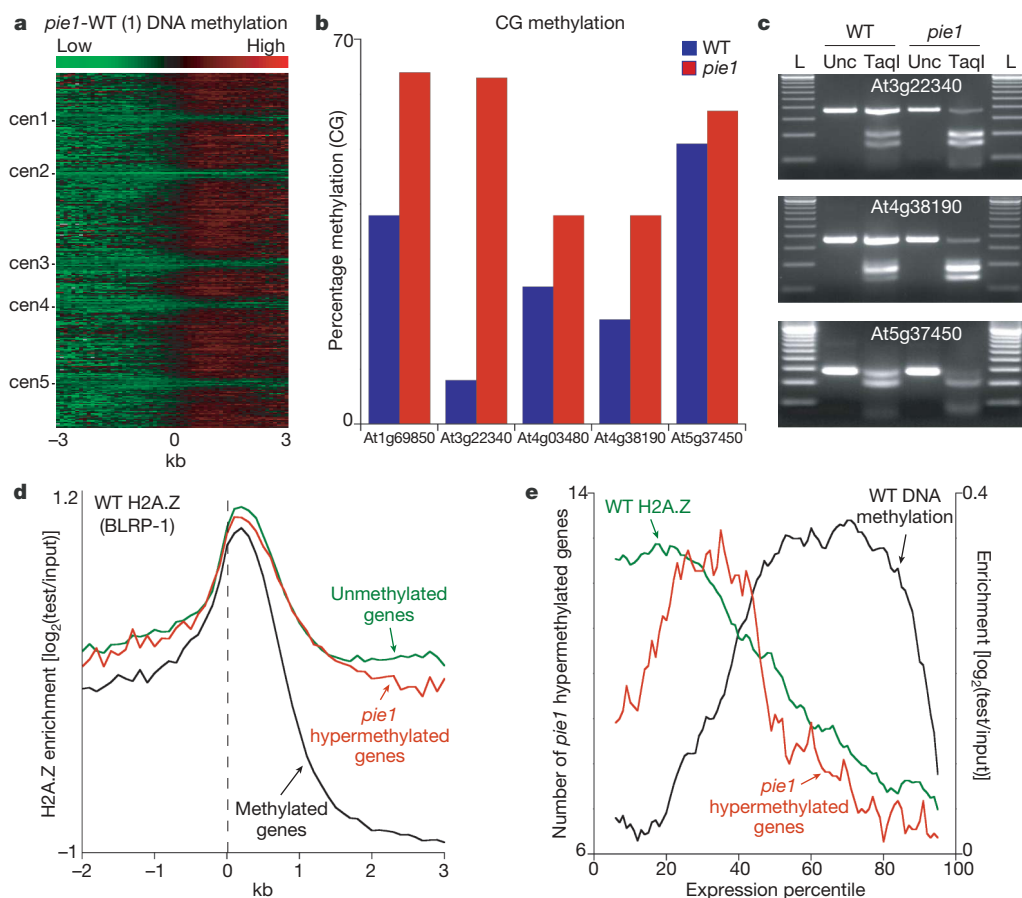

Figure 4 | H2A.Z protects from DNA methylation. a, All TAIR7 annotated sequences were aligned at the 5′ end and stacked from the top of chromosome 1 to the bottom of chromosome 5. The WT methylation pattern subtracted from the *pie1* methylation pattern is displayed as a heat map. Shown are results from data set 1; all three data sets are shown in Supplementary Fig. 16. **b,** Bisulphite sequencing results for five loci. We sequenced 12 clones from each genotype, except for At1g69850 (10 clones in *pie1*) and At4g38190 (11 clones in *pie1*). **c,** Polymerase chain reaction (PCR) products from bisulphite-converted genomic DNA were digested with TaqI, which recognizes TCGA and will cut only if the C is unconverted (and therefore methylated). L, 100 bp ladder; TaqI, PCR product digested with TaqI; Unc, uncut PCR product. Note the greater digestion, which represents greater methylation, in *pie1* compared to that in WT. **d,** All genes were

aligned at the 5′ end and average scores for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 3 kb into the gene (positive numbers). The data were smoothed with a 5-point sliding window. The dashed line represents the point of alignment. The black line traces methylated genes; the green line traces unmethylated genes and the red line traces genes hypermethylated in *pie1*. **e,** Genes were grouped into percentiles on the basis of transcription levels. The red line traces the number of genes hypermethylated in *pie1* within each percentile (left y axis). The black line traces DNA methylation enrichment (all genes) and the green line traces H2A.Z enrichment in unmethylated genes (right y axis). The data were smoothed with a 10-point sliding window. The scale of the right y axis was set to start at zero to enable comparison between methylation and H2A.Z.

around the 30th transcription percentile (Fig. 4e). This pattern is very different from that of normally methylated genes, which are most prevalent around the 70th percentile (Fig. 4e), and is also unlike unmethylated genes, which are enriched in both low and highly expressed genes[22]. *pie1* hypermethylated genes do, however, closely parallel the overall distribution of H2A.Z (Fig. 4e). These loci also include 17 of the 49 transposons that are enriched in H2A.Z and unmethylated in wild type (Supplementary Tables 3 and 6), a 10-fold overrepresentation ($P = 10^{-4}$, Fisher's exact test). Thus, sequences that are generally preferred targets of DNA methylation (gene bodies and transposons) are hypermethylated in *pie1*, consistent with the presence of low levels of DNA methylation in these sequences in wild type (Fig. 4b, c). The high levels of H2A.Z found at these loci apparently protect them from developing full-blown DNA methylation, probably explaining the observed relationship between gene transcription and DNA methylation[22].

How methylation silences genes has been a vexing question for decades. A popular model is that proteins that bind to methylated DNA engender silencing by recruiting histone deacetylases[6]. However, careful gene disruption studies in mice have shown that these proteins are unlikely to fully account for methylation-induced repression[29,30]. Previous work has provided strong evidence that H2A.Z contributes to promoter competence[16–19]. Therefore, exclusion of H2A.Z would represent a new mechanism of gene silencing by DNA methylation. H2A.Z incorporation, in turn, is likely to protect gene promoters from DNA methylation, contributing to gene activity and preventing silencing. Given that DNA methylation and H2A.Z are both ancient chromatin components, their interaction probably has an important general role in regulating eukaryotic gene expression.

METHODS SUMMARY

We adapted the biotin-mediated affinity purification system we developed in *Drosophila* tissue culture cells[21] to allow protein purification from *Arabidopsis* plants. Biotinylated H2A.Z was purified largely as described[21]. Endogenous H2A.Z was immunopurified (IP) as described[17], except the IP was performed in TNE (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA).

Our methylated DNA IP protocol (MeDIP), microarray design and labelling protocol are described in ref. 22. All labelled samples were sent to NimbleGen Systems for hybridization, except the *pie1* samples, which were hybridized at the Fred Hutchinson Cancer Research Center DNA array facility. For bisulphite sequencing, 2 μg genomic DNA for each sample was bisulphite-converted with the Qiagen EpiTect kit.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nature Struct. Biol.* **10**, 882–891 (2003).

2. Bernstein, E. & Hake, S. B. The nucleosome: a little variation goes a long way. *Biochem. Cell Biol.* **84**, 505–517 (2006).

3. Bhaumik, S. R., Smith, E. & Shilatifard, A. Covalent modifications of histones during development and disease pathogenesis. *Nature Struct. Mol. Biol.* **14**, 1008–1016 (2007).

4. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).

5. Gehring, M. & Henikoff, S. DNA methylation dynamics in plant genomes. *Biochim. Biophys. Acta* **1769**, 276–286 (2007).

6. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).

7. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nature Rev. Genet.* **7**, 21–33 (2006).

8. Mizuguchi, G. *et al.* ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).

9. Guillemette, B. & Gaudreau, L. Reuniting the contrasting functions of H2A.Z. *Biochem. Cell Biol.* **84**, 528–535 (2006).

10. Guillemette, B. *et al.* Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.* **3**, e384 (2005).

11. Li, B. *et al.* Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl Acad. Sci. USA* **102**, 18385–18390 (2005).

12. Millar, C. B., Xu, F., Zhang, K. & Grunstein, M. Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev.* **20**, 711–722 (2006).

13. Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).

14. Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**, 219–231 (2005).

15. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

16. Brickner, D. G. *et al.* H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol.* **5**, e81 (2007).

17. Deal, R. B., Topp, C. N., McKinney, E. C. & Meagher, R. B. Repression of flowering in *Arabidopsis* requires activation of *FLOWERING LOCUS C* expression by the histone variant H2A.Z. *Plant Cell* **19**, 74–83 (2007).

18. Meneghini, M. D., Wu, M. & Madhani, H. D. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**, 725–736 (2003).

19. Updike, D. L. & Mango, S. E. Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS Genet.* **2**, e161 (2006).

20. Venkatasubrahmanyam, S., Hwang, W. W., Meneghini, M. D., Tong, A. H. & Madhani, H. D. Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. *Proc. Natl Acad. Sci. USA* **104**, 16609–16614 (2007).

21. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genet.* **37**, 1090–1097 (2005).

22. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39**, 61–69 (2007).

23. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).

24. Vaughn, M. W. *et al.* Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5**, e174 (2007).

25. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).

26. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).

27. Xiao, W. *et al.* Imprinting of the *MEA* Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901 (2003).

28. Buck, M. J., Nobel, A. B. & Lieb, J. D. ChIPOTle: a user-friendly tool for the analysis of ChIP–chip data. *Genome Biol.* **6**, R97 (2005).

29. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genet.* **27**, 322–326 (2001).

30. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15**, 710–723 (2001).

METHODS

Transgenic lines. We adapted the biotin-mediated affinity purification system that we developed in *Drosophila* tissue culture cells[21] to allow protein purification from *Arabidopsis* plants. We constructed a binary plasmid that contained the *E. coli* biotin ligase, BirA, driven by the *Arabidopsis ACT2* promoter, and the *Arabidopsis* H2A.Z gene At1g52740 driven by its endogenous promoter and tagged at the amino terminus with the BLRP. BLRP is a high affinity substrate for BirA, which biotinylates a lysine residue within the peptide. We sent the plasmid to the UC Riverside Plant Transformation Research Center (http://www.ptrc.ucr.edu), where transgenic *Arabidopsis* lines were created by vacuum infiltration in ecotype Columbia.

Affinity purification. About 100 seeds were sterilized in 20% bleach and 0.5% Tween-20 for 10 min. Seeds were germinated in 300 ml of Gamborg's B-5 medium supplemented with 5 mM biotin, and roots harvested after four weeks. Four grams of roots were ground in liquid nitrogen to a fine powder, suspended in 20 ml of modified Honda buffer (25 mM Tris, pH 7.6, 0.44 M sucrose, 10 mM $MgCl_2$, 2 mM spermine, 0.1% Triton X-100, 10 mM β-mercaptoethanol) and homogenized with a tissue homogenizer. The homogenate was filtered through Miracloth, transferred to a 30 ml round-bottom glass tube, and spun at 4,000 r.p.m. (2,000*g*) at 4 °C in an SS-34 rotor for 10 min. The pellet was resuspended in Honda buffer B (Honda buffer minus spermine), spun in a microcentrifuge at 1,500 r.p.m.(200*g*) at 4 °C for 2 min, and resuspended in 1 ml of TNE (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA). The suspension was warmed to 37 °C and digested with micrococcal nuclease in the presence of 4 mM $CaCl_2$ (Supplementary Fig. 1) to liberate nucleosomes. The reaction was stopped with 25 mM EDTA and spun at high speed in a microcentrifuge for 5 min at 4 °C. Biotinylated proteins were purified from the supernatant as described[21]. Endogenous H2A.Z was immunopurified as described[17], except the IP was performed in TNE. The antibodies are predicted to cross-react with all three *Arabidopsis* H2A.Z proteins[17].

Microarray analysis. Our microarray design is described in ref. 22. We analysed DNA methylation in five independent samples from the Columbia ecotype: two from wild-type roots (WT root-1 and root-2), two from *met1-6* roots (*met1-6* root-1 and root-2), and one from *met1-6* aerial tissues. We followed our

protocol, as described in ref. 22, except we omitted the T7 RNA polymerase-mediated amplification step for all samples except aerial *met1-6*. Instead, sufficient amplification was achieved in the labelling step. We also used our wild-type aerial methylation data published in ref. 22.

For *pie1* methylation analysis, we mapped methylation in three *pie1* replicates and three matched wild-type controls (F_2 siblings). DNA was extracted from tissue collected from >100 whole 12-day-old seedlings to eliminate the possibility of detecting random variations in DNA methylation[24]. The samples were amplified with the Sigma WGA2 kit before labelling. For ChIPOTle analysis[28], outliers were removed from each data set by median smoothing (3-probe window), the three *pie1* and wild-type data sets were averaged, wild type was subtracted from *pie1*, and the resulting data set was smoothed (triangular smoothing, $y = 0.25(x_{n-1}) + 0.5(x_n) + 0.25(x_{n+1})$) and normalized to a mean of zero. We removed the 270-kb mitochondrial DNA insertion on chromosome 2 before analysis. 1,201 peaks were called with a conservative threshold of $P < 10^{-7}$. As a control, we determined the number of 'negative' peaks that would represent hypomethylation: only 53 peaks were called. Even assuming the unlikely scenario that all the negative peaks are false positives, the false-positive rate would be 4%.

We assayed H2A.Z in eight samples: BLRP-1 and BLRP-2 were from one transgenic line, BLRP-3 and BLRP-4 from an independent transgenic line, wild-type Ab-1 and *met1-6* Ab-1 were paired immunoprecipitation experiments from wild-type and *met1-6* roots, respectively, and wild-type Ab-2 and *met1-6* Ab-2 were a second set of paired experiments. All samples except BLRP-2, BLRP-3 and BLRP-4 were amplified by T7 RNA polymerase[22]; the rest were sufficiently amplified in the labelling step.

Expression analysis of two independent *met1-6* RNA samples (paired with two independent wild-type samples) was carried out as described in ref. 22, except random hexamers were used for complementary DNA synthesis instead of an oligo d(T) primer. All labelled samples were sent to NimbleGen Systems for hybridization, except the *pie1* samples, which were hybridized at the FHCRC DNA array facility.

Bisulphite sequencing. Two micrograms of genomic DNA for each sample were bisulphite-converted with the Qiagen EpiTect kit. PCR products were cloned with the Invitrogen PCR4 TOPO kit. Primer sequences are available on request.

LETTERS

Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs

Kouji Hirota[1,2]†, Tomoichiro Miyoshi[2], Kazuto Kugou[1,2], Charles S. Hoffman[3], Takehiko Shibata[1] & Kunihiro Ohta[1,2]

Recent transcriptome analyses using high-density tiling arrays[1–3] and data from large-scale analyses of full-length complementary DNA libraries by the FANTOM3 consortium[4,5] demonstrate that many transcripts are non-coding RNAs (ncRNAs). These transcriptome analyses indicate that many of the non-coding regions, previously thought to be functionally inert, are actually transcriptionally active regions with various features. Furthermore, most relatively large (~several kilobases) polyadenylated messenger RNA transcripts are transcribed from regions harbouring little coding potential. However, the function of such ncRNAs is mostly unknown and has been a matter of debate[2]. Here we show that RNA polymerase II (RNAPII) transcription of ncRNAs is required for chromatin remodelling at the fission yeast *Schizosaccharomyces pombe fbp1*[+] locus during transcriptional activation. The chromatin at *fbp1*[+] is progressively converted to an open configuration, as several species of ncRNAs are transcribed through *fbp1*[+]. This is coupled with the translocation of RNAPII through the region upstream of the eventual *fbp1*[+] transcriptional start site. Insertion of a transcription terminator into this upstream region abolishes both the cascade of transcription of ncRNAs and the progressive chromatin alteration. Our results demonstrate that transcription through the promoter region is required to make DNA sequences accessible to transcriptional activators and to RNAPII.

In the fission yeast *S. pombe*, transcriptional regulation of *fbp1*[+], which is robustly induced by glucose starvation, has been well studied to identify the signal transduction pathways and the transcriptional regulators involved[6,7]. We have detected long and rare *fbp1*[+] transcripts that are transiently expressed during starvation-induced derepression. As shown in Fig. 1, we analysed *fbp1*[+] transcription in the course of glucose starvation and observed at least four distinct species of *fbp1*[+] transcripts (Fig. 1A, indicated by arrowheads a–d). The main transcript (labelled d) represents the functional *fbp1*[+] transcript and appears after 60 min of glucose starvation. Each of the longer transcripts (a, b and c) appear at 0–20, 10–60 and 20–60 min after glucose starvation, respectively, suggesting that such long and rare transcripts initiate upstream of the main transcriptional initiation site (Fig. 1A). To test this possibility, we used probes covering UAS1 or UAS2, both of which are essential *cis*-elements for *fbp1*[+] induction[8], to detect these long transcripts. The long transcripts (a and b, or a, b and c) were detected by the UAS1 or UAS2 probes, respectively, indicating that transcripts a, b and c initiate in the region around UAS1 and UAS2 (Fig. 1B and C). Moreover, northern analysis using strand-specific probes showed that all *fbp1*[+] transcripts are sense RNAs, indicating that this is not an antisense RNA mechanism (Fig. 1D). To determine the actual transcription initiation sites of these long RNAs, we performed a rapid amplification of cDNA ends (RACE) analysis. This study demonstrated that the long transcripts initiate upstream from UAS1

(transcript a), within UAS1 (transcript b) and in the region between UAS1 and UAS2 (transcript c; Fig. 1E). In addition, these RNAs are polyadenylated, as we used a poly-T primer in the construction of the RACE library. Sequence analysis indicates that there is no splicing of these RNAs. Notably, some of the long RNA are polyadenylated at the site within the *fbp1*[+] coding region (Supplementary Fig. 1) and do not produce a protein product (Supplementary Fig. 2). Thus, these long precursor transcripts do not seem to have any special protein-encoding function.

We next examined the RNAPII binding around the *fbp1*[+] locus using chromatin immunoprecipitation (ChIP) analysis. For the quantitative analysis, the *fbp1*[+] locus was divided into ~250 base pair (bp) segments (probes I, II and III, as shown in Fig. 2a), and the probes for each region were used in a slot-blot analysis to measure ChIP efficiency within those segments. We detected increased binding of RNAPII in regions I, II and III at 0–30, 10–30 and 30–180 min after glucose starvation, respectively, indicating that the association of RNAPII to the DNA shifts 5′ to 3′ along the *fbp1*[+] control region towards the promoter (Fig. 2a, b).

Given the changes in the length of the transcripts, as well as the alteration of RNAPII distribution, we examined whether the chromatin configuration changes in a similar manner, because transcriptional activity correlates with an open chromatin configuration[9]. The chromatin in UAS1 is protected from micrococcal nuclease (MNase) digestion before glucose starvation, although a couple of intense bands appear around UAS1 (Fig. 2c, grey arrowheads). After glucose starvation, intense bands appear in UAS1 within 10 min (Fig. 2c, arrowheads). The intensity of the bands located between UAS1 and UAS2 increases within 10–30 min after glucose starvation (Fig. 2c, dotted line). Furthermore, the bands around the TATA box gradually intensify after 30 min or more of glucose starvation (Fig. 2c, thick line). These results indicate that chromatin remodelling is initiated far upstream from the *fbp1*[+] promoter and is induced in a stepwise manner 5′ to 3′ towards the *fbp1*[+] promoter. Moreover, such chromatin alterations coincide with changes to the transcriptionally active start sites (compare Fig. 1A and Fig. 2c).

Having established a strong correlation between chromatin remodelling and transcriptional initiation events, we next examined whether transcription from these upstream sites is required for chromatin remodelling in the *fbp1*[+] promoter. We inserted a transcriptional terminator between UAS2 and the TATA box to prevent the passage of RNAPII through the promoter region. As shown in Fig. 3a, insertion of the *nmt*-terminator 3′ to UAS2 (EcoT22I site) eliminates *fbp1*[+] induction. When using a probe that covers UAS2 (upstream of the inserted terminator), we detected a short transcript reflecting premature termination of an upstream transcript (Fig. 3a). These results indicate that the passage of RNAPII is blocked by the terminator sequence. In this mutant, chromatin remodelling events are

[1]Cellular & Molecular Biology Laboratory, RIKEN Advanced Science Institute, Wako-shi, Saitama 351-0198, Japan. [2]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan. [3]Biology Department, Boston College, Chestnut Hill, Massachusetts 02467, USA. †Present address: Department of Radiation Genetics, Kyoto University Graduate School of Medicine, Yoshida Konoe, Sakyo-ku, Kyoto 606-8501, Japan.

unaffected at UAS1 and UAS2 (Fig. 3b, indicated by arrowhead and dotted line, respectively) but do not appear downstream of the terminator sequence. In the control experiment, insertion of the sequence between UAS1 and UAS2 (−1160 to −534 bp from the ATG), which apparently has no termination activity, did not prevent chromatin remodelling downstream of the inserted sequences (Fig. 3d, thick line). Thus, it is probable that the passage of RNAPII through this region is pivotal for the chromatin remodelling events. Quantification of MNase sensitivity around the TATA box also supports this notion (Supplementary Fig. 3). Conversely, derepression of $fbp1^+$ is greatly reduced in this mutant, indicating that the distance between UAS2 and the TATA box is important for the level

of transcriptional activation of $fbp1^+$ (Fig. 3c). This also indicates that chromatin remodelling is not a simple consequence of the massive transcriptional activation. Furthermore, the insertion of the terminator sequence 5′ to UAS2 (NaeI site), in which the distance between UAS2 and the TATA box is not altered, also prevents both derepression of $fbp1^+$ and chromatin remodelling (Fig. 3e, f). A functional role for transcription of these ncRNAs in $fbp1^+$ derepression is also demonstrated by the effect of the RNAPII inhibitor phenanthroline[10], which inhibits glucose-starvation-induced chromatin remodelling (Supplementary Fig. 4). These results indicate that the passage of RNAPII through the $fbp1^+$ upstream region is vital for chromatin remodelling and transcriptional derepression of $fbp1^+$.
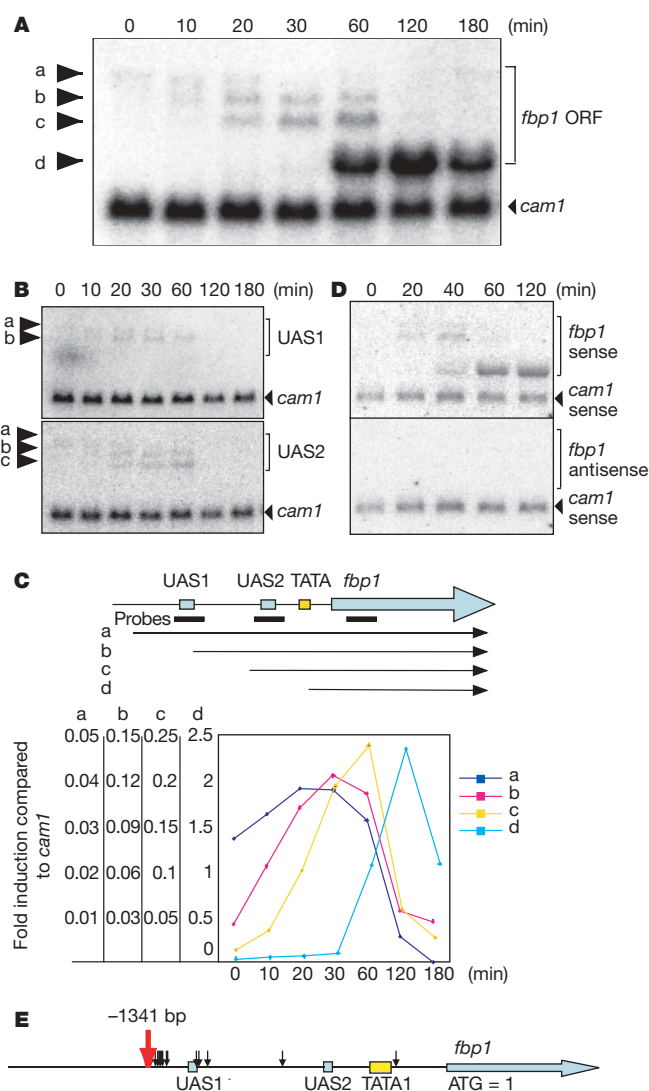

Figure 1 | Long and rare $fbp1^+$ transcripts during transcriptional activation. Wild-type cells (K131) were grown to mid-log phase (1×10^7 cells ml^{-1}) in YER (yeast extract repression) medium containing 6% glucose and were transferred to YED (yeast extract derepression) medium containing 0.1% glucose and 3% glycerol. Cells were collected at the indicated time. **A**, Variable length $fbp1^+$ transcripts (labelled as a–d). The $fbp1^+$ transcripts were analysed by northern analysis using the $fbp1$ ORF probe. The $cam1^+$ (ref. 24) transcript was used as an internal control. **B**, Long and rare transcripts initiate from the 5′ region of $fbp1^+$. The probes covering UAS1 and UAS2 were used. **C**, Quantification of the results in **A**. The vertical axes (a, b, c and d) show the fold induction of each transcript normalized to $cam1^+$. **D**, All $fbp1^+$ transcripts in **A** are sense RNA. The $fbp1^+$ transcripts were detected using strand-specific probes. **E**, Transcription initiation sites of $fbp1^+$ were determined by 5′-RACE. The red and black arrows denote initiation sites of the longest RNA and other RNAs, respectively. The 'A' of the first 'ATG' is represented as 1.
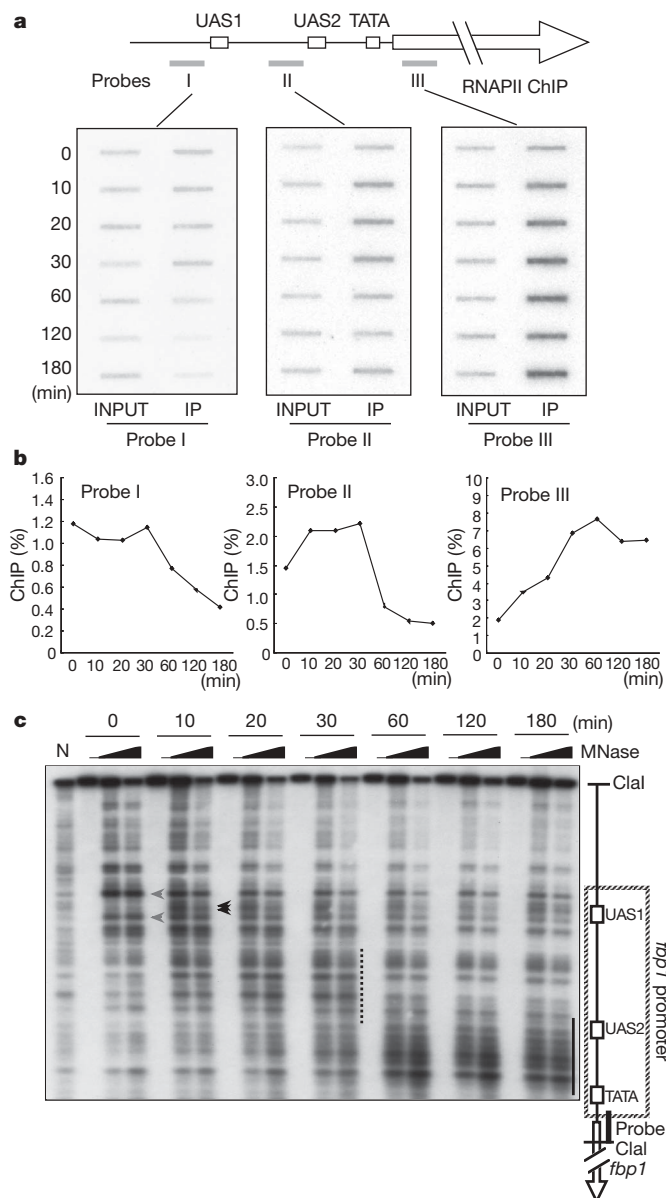
Figure 2 | The RNAPII binding sites shift from the 5′ to 3′ region in the $fbp1^+$ promoter, coupled with chromatin remodelling. a, Distribution of RNAPII was analysed by ChIP. The input and chromatin immunoprecipitated (IP) DNA were measured by slot blotting and hybridization with the probes as indicated. The numbers represent the minutes after glucose starvation. **b**, Quantification of the RNAPII binding. The vertical and horizontal axes represent ChIP efficiency (expressed as a percentage) and the minutes after glucose starvation. Similar results were obtained from repeated experiments. **c**, Cascade of chromatin alteration in the $fbp1^+$ promoter. Chromatin was analysed using the cell culture of Fig. 1A. The open arrow denotes the $fbp1^+$ coding region. Lane N represents partial digestion of naked DNA with MNase.
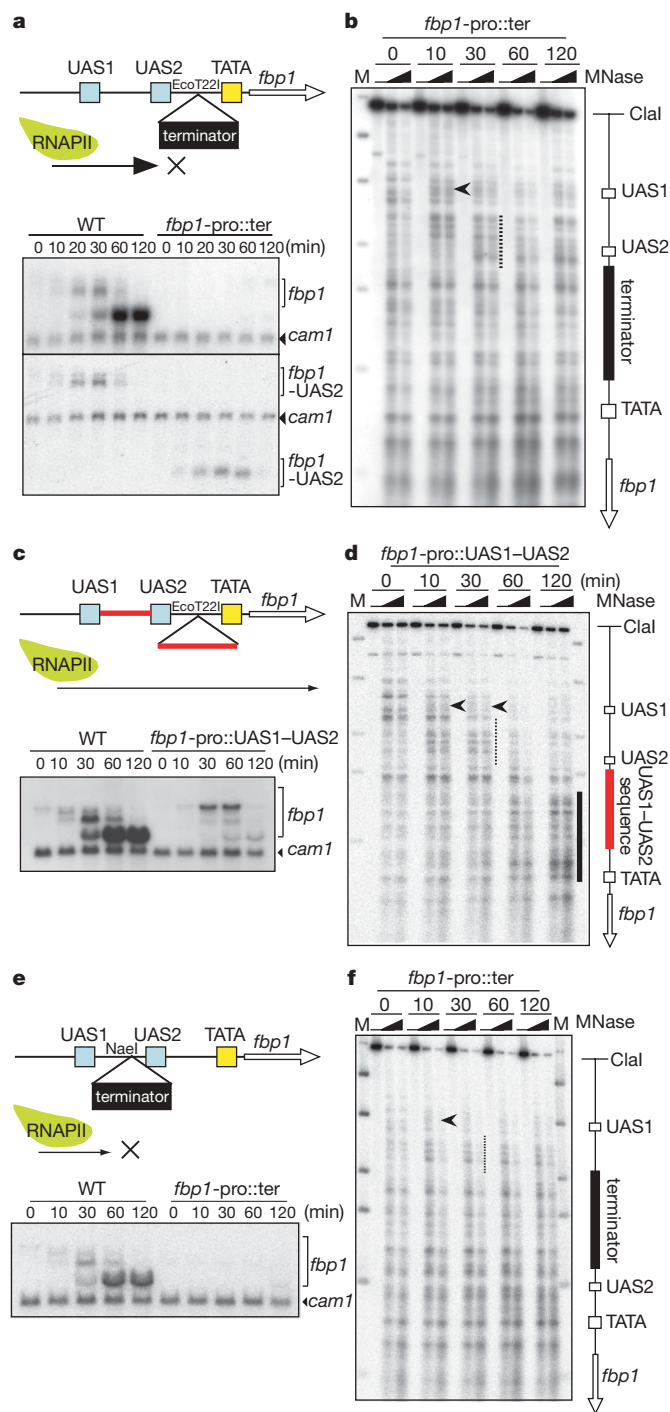
Figure 3 | RNAPII passage along the *fbp1*⁺ promoter is required for chromatin remodelling. a, Wild-type (WT; K176) and *nmt*-terminator-inserted mutant (EcoT22I site; PKH522) cells were cultured as in Fig. 1. The *fbp1*⁺ transcripts were analysed by northern analysis using probes covering the *fbp1*⁺ ORF or UAS2. b, The chromatin structure of *fbp1*⁺ in the cells from the culture in **a**. c, d, Insertion of a control sequence without a terminator (the sequence between UAS1 and UAS2) at EcoT22I site (PKH575 strain) does not prevent chromatin remodelling. e, f, The *fbp1*⁺ transcripts and chromatin structure in the control strain with an *nmt*-terminator insertion 5′ to UAS2 (NaeI site; PKH558). Note that this insertion does not affect the distance between UAS2 and TATA. Lane M represents the molecular weight marker λ-EcoT14I (Takara). *fbp1*-pro::ter and *fbp1*-pro::UAS1–UAS2 represent the insertion of the terminator sequence and the sequence between UAS1 and UAS2, respectively, into the *fbp1*⁺ promoter.

Previous studies have shown that the transcriptional activators Atf1–Pcr1 (a heterodimeric basic leucine zipper protein) and Rst2 (a zinc-finger protein) bind to UAS1 and UAS2, respectively, and that the Tup11 and Tup12 co-repressors bind to a broad region from UAS1 to UAS2 (refs 8, 11 and 12). Thus, these factors seem to regulate the chromatin configuration in collaboration with RNAPII, thereby regulating *fbp1*⁺ transcription. To test this assumption, we analysed *fbp1*⁺ transcription in mutants lacking these factors. In the *atf1*⁻ mutant, transcripts a and b are expressed normally, whereas transcripts c and d are absent (Fig. 4A). In the *rst2*⁻ mutant, transcripts a, b, and c are expressed normally, whereas transcript d is absent (Fig. 4A). Moreover, *fbp1*⁺ derepression is recovered by deleting both *tup11*⁺ and *tup12*⁺, indicating that Atf1 and Rst2 are dispensable for *fbp1*⁺ induction in the absence of both Tup proteins. Expression of transcript c is not restored in the *atf1*⁻ *tup11*⁻ *tup12*⁻ mutant, suggesting that Atf1 is essential to induce transcript c. The requirement of transcript c for *fbp1*⁺ derepression is bypassed by deleting the *tup11*⁺ and *tup12*⁺ genes. Chromatin remodelling events and RNAPII loading around the TATA box are severely impaired in an *atf1*⁻ mutant, demonstrating that the progression of ncRNA initiation events mediated by Atf1 is essential to convert chromatin to an RNAPII accessible state (Fig. 4B). The passage of RNAPII through UAS1 may not be important for Atf1 loading at UAS1, because we observe an almost wild-type level of the *fbp1*⁺ transcripts c and d, which require Atf1 binding to UAS1, even when the passage of RNAPII through UAS1 is abolished by a terminator insertion upstream of UAS1 (Supplementary Fig. 5). In contrast, ChIP data indicate that the passage of RNAPII through UAS2 is pivotal for the recruitment of Rst2 to UAS2 (Fig. 4D).

These results lead us to propose a model for the regulation of the chromatin structure at *fbp1*⁺ (Fig. 4E). Under glucose-rich conditions, RNAPII initiates rare transcripts from a site far upstream from the authentic *fbp1*⁺ promoter, but does not initiate the robust activation of *fbp1*⁺ transcription at the promoter owing to the Tup-dependent repressive chromatin structure (Fig. 4E, (i)). After glucose starvation, Atf1 first binds to UAS1, allowing RNAPII to initiate transcription at sites b and c, to overcome the repressive role of the Tup proteins (Fig. 4E, (ii)). The second step of the transcriptional cascade disrupts the chromatin structure within the *fbp1*⁺ promoter, thereby allowing Rst2 to bind to UAS2 (Fig. 4E, (iii)). The chromatin structure around the TATA box then becomes open and suitable for the stable binding of the basic transcriptional machinery to activate transcription of the protein-encoding message fully (Fig. 4E, (iv)). We observed that the chromatin structure around the TATA box is converted into a partially accessible state in the *rst2*⁻ mutant (Fig. 4B). In addition, the binding of RNAPII around the TATA box is partially reduced and is unstable in the *rst2*⁻ mutant (Fig. 4C). Thus, it is plausible that Rst2 and the Tup proteins together regulate the stability of RNAPII binding at the TATA box ((iv) in Fig. 4E). Alternatively, it is possible that transcribed ncRNA molecules function *in trans* to induce chromatin remodelling. However, this is unlikely, as expression of the ncRNA *in trans* from a plasmid in the terminator-inserted mutant does not suppress the defect in chromatin remodelling around the TATA box (data not shown).

In this study, we provide evidence for chromatin remodelling induced by RNAPII transcription of precursor ncRNAs. Related to this observation, we recently demonstrated a shift of the transcriptional initiation site coupled to histone acetylation and chromatin remodelling in *ade6-M26*, which, like *fbp1*⁺, is regulated by Atf1 and the Tup proteins[13]. Moreover, our genome tiling array analysis of *S. pombe* transcripts uncovered further loci (some with potential Rst2 binding sites) in which the transcription events occur far upstream from promoters after glucose starvation (Supplementary Fig. 6). In *Saccharomyces cerevisiae*, a previous study reported a role for transcription upstream of the *SER3* gene in *SER3* repression[14], whereas antisense transcription at the *PHO5* promoter regulates the rate of
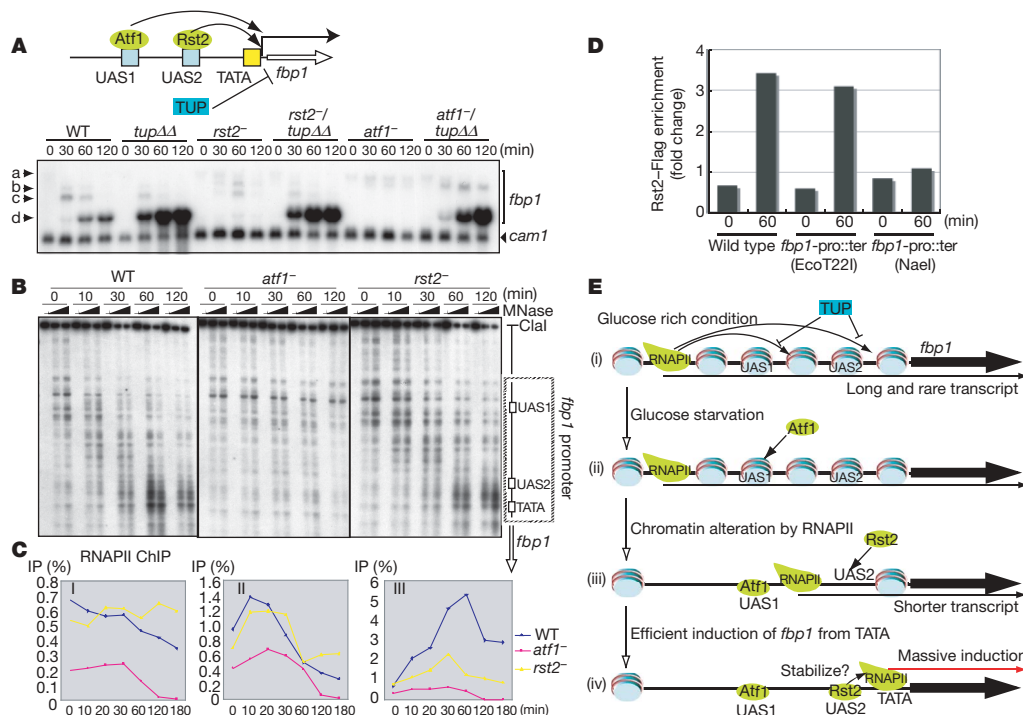
Figure 4 | Regulation by Atf1, Rst2 and the Tup co-repressors. A, The *fbp1*[+] transcripts in wild-type (K131), *tup11*[−]/*tup12*[−] (*tupΔΔ*; PKH40), *rst2*[−] (PKH108), *rst2*[−]/*tupΔΔ* (PKH107), *atf1*[−] (PKH64) and *atf1*[−]/*tupΔΔ* (PKH193) cells. **B**, **C**, Chromatin alteration and RNAPII binding around TATA was severely and partially impaired in *atf1*[−] and *rst2*[−], respectively. Chromatin (**B**) and the binding of RNAPII (**C**) in wild-type (K131), *atf1*[−] (PKH64) and *rst2*[−] (PKH108) cells. IP, immunoprecipitation. **D**, The

passage of RNAPII through UAS2 is vital for Rst2 binding to UAS2. The *rst2-flag* strain[12] carrying the *nmt*-terminator 5′ or 3′ to UAS2 was glucose starved for 60 min, and anti-Flag ChIP was performed using UAS2 and *lys1*[+] primer sets. Relative enrichment of the UAS2 sequence in precipitates as compared to the *lys1*[+] sequence was determined by quantitative PCR. **E**, A model of chromatin remodelling by a cascade of ncRNAs.

histone eviction and gene activation[15]. This current study supports the notion raised by these previous articles that non-canonical transcripts can regulate gene expression, and provides further evidence that an RNAPII-transcribed cascade of ncRNAs can act as 'pioneers' to disrupt chromatin. Similar to the present study, ncRNA transcription in the *Drosophila* bithorax complex regulates homeotic gene activation[16,17]. However, in this case the *Drosophila* ncRNAs can activate transcription when expressed *in trans* to the target gene[17]. Moreover, the presence of long and rare intergenic transcripts in the human β-globin locus, which contains an open chromatin structure with hyperacetylated histones[18–20], suggests that transcription-mediated regulation of gene expression could also exist in humans. Furthermore, recent transcriptome studies identifying many alternative RNA start sites in yeast genes[3,21] also support the idea of transcription-mediated gene expression. The present study demonstrates a new mechanistic role for ncRNAs in gene activation of eukaryotes.

METHODS SUMMARY

All media and growth conditions were as described[22]. The fission yeast strains used in this study are listed in Supplementary Table 1. ChIP was performed as described previously using anti-RNA polymerase II antibody (clone CTD4H8; Upstate) or anti-Flag M2 antibody (Sigma)[12]. DNA concentrations were quantified using fast real-time PCR system 7300 (Applied Biosystems) and SYBR premix EX TaqII (Takara) using the following primer set: UAS2, AATTGCAGTATGTCATTTGTTTAGCAG and ACTACAGGGCAATGCTG TTTCA; *lys1*[+], TCCAGGACGCACAACAACAC and GTAAGGCAAGCGGGA TTACG. Isolation and MNase digestion of chromatin fractions were performed as described previously[23]. The probe used for the chromatin analysis is described previously[22]. The strand-specific probes for northern analysis were prepared by 5′ end-labelling oligonucleotides with ^{32}P using a MEGALABEL 5′ end-labelling kit (Takara). 5′-RACE was carried out using a SMART RACE cDNA amplification kit (Clontech). The 5′ ends of transcripts were amplified by PCR using the universal primer mix included in the kit and the gene-specific primer GCTGGCCAGGTGAATCCTAGCTGAA. PCR products were gel-purified

(QIAquick; Qiagen) and cloned into pCR2.1TOPO (Invitrogen). The sequences were determined using the M13 primer. Transcriptome analysis was performed with genome tiling array (Affymetrix GeneChip *S. pombe* tiling 1.0FR array). To construct terminator-inserted mutants, a 1.0-kb fragment carrying the *nmt*-terminator sequence from a vector pREP1was inserted into the EcoT22I (for PKH522) or NaeI (for PKH558) site of the *fbp1*[+] promoter. To construct the control strain PKH575, the sequence covering −1160 to −534 bp relative to the *fbp1*[+] open reading frame (ORF) was inserted into the EcoT22I site of the *fbp1*[+] promoter.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
2. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
3. Wilhelm, B. T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243 (2008).
4. The FANTOM Consortium. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
5. Hayashizaki, Y. & Carninci, P. Genome Network and FANTOM3: assessing the complexity of the transcriptome. *PLoS Genet.* **2**, e63 (2006).
6. Hoffman, C. S. Glucose sensing via the protein kinase A pathway in *Schizosaccharomyces pombe*. *Biochem. Soc. Trans.* **33**, 257–260 (2005).
7. Janoo, R. T., Neely, L. A., Braun, B. R., Whitehall, S. K. & Hoffman, C. S. Transcriptional regulators of the *Schizosaccharomyces pombe fbp1* gene include two redundant Tup1p-like corepressors and the CCAAT binding factor activation complex. *Genetics* **157**, 1205–1215 (2001).
8. Neely, L. A. & Hoffman, C. S. Protein kinase A and mitogen-activated protein kinase pathways antagonistically regulate fission yeast *fbp1* transcription by employing different modes of action at two upstream activation sites. *Mol. Cell. Biol.* **20**, 6426–6434 (2000).
9. Wolffe, A. *Chromatin: Structure and Function* 3rd edn (Academic, 1997).
10. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D. & Hughes, T. R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays

reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.* **24**, 5534–5547 (2004).

11. Higuchi, T., Watanabe, Y. & Yamamoto, M. Protein kinase A regulates sexual development and gluconeogenesis through phosphorylation of the Zn finger transcriptional activator Rst2p in fission yeast. *Mol. Cell. Biol.* **22**, 1–11 (2002).

12. Hirota, K., Hoffman, C. S. & Ohta, K. Reciprocal nuclear shuttling of two antagonizing Zn finger proteins modulates Tup family corepressor function to repress chromatin remodeling. *Eukaryot. Cell* **5**, 1980–1989 (2006).

13. Hirota, K., Mizuno, K. I., Shibata, T. & Ohta, K. Distinct chromatin modulators regulate the formation of accessible and repressive chromatin at the fssion yeast recombination hotspot *ade6–M26*. *Mol. Biol. Cell* **19**, 1162–1173 (2008).

14. Martens, J. A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* **13**, 136–142 (2003).

15. Uhler, J. P., Hertel, C. & Svejstrup, J. Q. A role for noncoding transcription in activation of the yeast *PHO5* gene. *Proc. Natl Acad. Sci. USA* **104**, 8011–8016 (2007).

16. Schmitt, S., Prestel, M. & Paro, R. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.* **19**, 697–708 (2005).

17. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311**, 1118–1123 (2006).

18. Bender, M. A., Bulger, M., Close, J. & Groudine, M. β-*globin* gene switching and DNase I sensitivity of the endogenous β-*globin* locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).

19. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-*globin* locus. *Mol. Cell* **5**, 377–386 (2000).

20. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* **11**, 2494–2509 (1997).

21. Miura, F. *et al.* A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl Acad. Sci. USA* **103**, 17846–17851 (2006).

22. Hirota, K., Hoffman, C. S., Shibata, T. & Ohta, K. Fission yeast Tup1-like repressors repress chromatin remodeling at the *fbp1*⁺ promoter and the *ade6–M26* recombination hotspot. *Genetics* **165**, 505–515 (2003).

23. Hirota, K., Steiner, W. W., Shibata, T. & Ohta, K. Multiple modes of chromatin configuration at natural meiotic recombination hot spots in fission yeast. *Eukaryot. Cell* **6**, 2072–2080 (2007).

24. Takeda, T. & Yamamoto, M. Analysis and *in vivo* disruption of the gene coding for calmodulin in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **84**, 3580–3584 (1987).

Author Contributions K.H. and K.O. conceived the project. K.H. planned, performed and analysed most of the experiments. T.M. performed Rst2-ChIP and RACE experiments. T.M. and K.K. performed the genome tiling array experiment. K.H. wrote most of the paper. C.S.H., T.S. and K.O. gave scientific advice and financial support. K.O. supervised the project.

Author Information The raw and analysed data from the genome tiling array experiment are available from ArrayExpress (http://www.ebi.ac.uk/arrayexpress) under the accession numbers E-MEXP-1747. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.H. (khirota@rg.med.kyoto-u.ac.jp) or to K.O. (kohta@bio.c.u-tokyo.ac.jp).

nature

METHODS

Genome wide transcriptome analysis using genome tiling array. Total RNA was prepared from wild-type cells (K131, at 0, 15 and 60 min after transfer to YED-containing 0.1% glucose and 3% glycerol) for northern analysis. After purification of total RNA using DNase I (amplification grade) and the PureLink Micro-to-Midi Total RNA Purification system (Invitrogen), mRNA was isolated using FastTrack MAG mRNA Isolation kits (Invitrogen). cDNA synthesis, *in vitro* transcription with labelling, and fragmentation were performed with One-Cycle Target Labelling and Control Reagents (Affymetrix). Biotin-labelled cRNA was hybridized for 16 h at 45 °C on an Affymetrix GeneChip *S. pombe* Tiling 1.0FR array. The washing and staining, the scanning, and the data extraction were performed on Fluidics Station 400, GeneChip Scanner 3000 and Affymetrix GCOS 1.4, respectively. Normalization and calculation of the signal intensity at each probe position were carried out using Affymetrix Tiling Analysis Software v1.1 with a 500 bp window size (bandwidth), which roughly corresponds to half of the mean intergenic distance in *S. pombe*. To consider the individual hybridization behaviour of every probe, the signal of our RNA hybridization (our Affymetrix CEL file) was compared with that of genomic DNA hybridizations (Wilhelm's three DNA CEL files[3], downloaded from ArrayExpress under accession number E-MTAB-18) during the analysis (two-comparison analysis; our Affymetrix CEL file as the treatment group, and Wilhelm's DNA CEL files as the control group). To search for genes that were derepressed during glucose starvation, 15 min or 60 min (−) glucose CEL files were compared with 0 min (−) glucose CEL files.

naturejobs

A t what has been billed as the first global meeting to focus on research ethics in graduate education, representatives from the United States, Canada, Europe, China, Australia, Hong Kong and Botswana agreed on a set of principles designed to help suppress the amount of fabrication and plagiarism around the world. Although somewhat vague, the guidelines might at least help draw attention to the problem.

The September meeting, held in Florence, Italy, and run by the US Council of Graduate Schools, came up with some broad consensus points — "scholarly integrity is a core value of all universities", for example. It also proposed ways to consolidate the vast differences in ethical principles between nations: exchanging "best practices and resources" such as codes of conduct, regulatory frameworks and curricular materials. The meeting concluded with delegates drawing up five action points for organizations to consider when setting up collaborations. These included: developing an open-access website that could be used to exchange resources and best practices; using dual and joint degree programmes to standardize scholarly integrity; and finding ways to address ethical dilemmas that arise from the mobility of scholars — such as inconsistent rules related to plagiarism between countries.

Sceptics would point out that the details of the proposals are scant, broad-based and incomplete. And no mention was made of how the principles might be enforced to ensure compliance, especially on an international basis. But perhaps even just the discussion and deliberation of research ethics will help research leaders to recognize instances of misconduct and prepare them to address violations. Given the increasing mobility of scientists, universities should work to lessen the risk of some nations or regions becoming known as places with substandard research ethics. After all, if science is to be made rigorous throughout the world, then leaders should at least attempt to globalize a culture of responsibility before budding researchers earn permanent jobs and take those values into their own labs.

Gene Russo is editor of *Naturejobs*.

German mentors steer a new course

Nurturing independence is appreciated, as **Alison Abbott** finds out.

The stereotypical German professor is an intimidating man. He is stern and will brook no disobedience. Young scientists call him Herr Professor and depend on him for publications, money, meeting invitations and ideas — everything except independence itself.

This stereotype is, fortunately, outdated. And although the German academic system remains very hierarchical, many established German scientists do go out of their way to consider the next generation, helping them in the lab and providing the skills they will need when they leave. Recognizing the importance of these changes, and wishing to encourage them, *Nature* decided to hold its travelling mentor competition in Germany this year.

Three winners of the *Nature* Awards for Mentoring in Science were selected from candidates proposed by appreciative protégés who have established accomplished careers of their own. Physicist Klaas Bergmann, senior research professor at the Technical University of Kaiserslautern, shared the lifetime achievement award with molecular neuroscientist Heinrich Betz, director of the Max Planck Institute of Brain Research in Frankfurt. Bioinformatician Peer Bork, of the European Molecular Biology Laboratory in Heidelberg — and the most cited European researcher in molecular biology and genetics — won the mid-term career award. The jury also made special mention of Ania Muntau, a researcher in molecular medicine at the University of Munich who has done much to promote the careers of other female clinicians in Germany.

"It was a hard choice," says Ulrike Beisiegel, a molecular-cell biologist at the University of Hamburg and head of the awards jury. "But there was something extra special about the deep involvement of the winners with their mentees, going well beyond just scientific teaching and a communal ski trip."

> **"My motto: let the young guys work freely — trust their imagination and creativity."**
> — Heinrich Betz

Mentees report that the lab chiefs they nominated helped them learn how to navigate the world of science inside and outside their labs — dealing with scientist rivalries, handling university bureaucracy and steering through the complexities of government and ministry science policies. Their mentors helped them write their own grants, run their own projects and develop all the other practical skills needed for independence. And they provided the right balance of structure and independence, encouragement and instructive criticism.

Crucially, the winning mentors remained advisers long after their protégés left the lab, providing moral and practical support as new careers struggled for traction. "Good mentoring is an important but often overlooked skill, which seems to have a valuable component of heritability," says *Nature*'s editor-in-chief,

Philip Campbell, who initiated the awards in 2005. "Over the years of the prize, I've noticed that many former mentees now with their own careers say that they try to provide the same style of mentoring as they received from the mentor they nominated."

Bergmann says he is most proud of the fact that protégés of his who returned to their home institutes in eastern Europe have acted as role-model scientists and administrators there.

Nikolay Vitanov, now a professor at the University of Sofia in Bulgaria, credits Bergmann with helping him establish a scientific career in his native country. "I now have an active and growing young research group in Sofia, a model for others to follow," says Vitanov. Aigars Eker, now a lab chief at the University of Latvia, was able to establish a lab and laser centre thanks to Bergmann's gift of scientific equipment. Aram Papoyan, now a director at the Institute for Physical Research of the National Academy of Sciences of Armenia in Ashtarak, lauds Bergmann's detail-oriented approach to science. "I have adopted his working methods and style," he says.

Heinrich Betz believes his biggest contribution has been giving his students the leeway to achieve their own successes and suffer their own failures. He allows "any experiment they consider essential, providing that it is affordable", although he debates every proposal. "Mostly I am correct, but sometimes they are right and this has not infrequently been rewarded by serendipitous findings," says Betz. "My motto: let the young guys work freely — trust their imagination and creativity." Former protégé Dieter Langosch, now a professor at Technical University Munich, adds: "He has a warm character and knows how to motivate people — but he does not hesitate to set appropriate limits."

Scientists mentored by Bork likewise praised his convivial teaching style. "He conveys an atmosphere of trust and confidence and has a playful style of thinking and working," says Christian von Mering, an associate professor at the University of Zurich and group leader at the Swiss Institute of Bioinformatics. Francesca Ciccarelli, now an assistant professor at the European Institute of Oncology, praised Bork's 'sixth sense' for science discovery. "He can smell where the real biological signal is," she says. "I learnt from him both scientifically and personally."

Bork counsels his protégés to strive for challenging projects and resist settling on mediocre job offers. "I convinced them to reject them, to believe in themselves and go for better ones," he says. "It always worked." ■

Alison Abbott is *Nature*'s senior European correspondent.

> **"There was something about the involvement of the winners with their mentees that went well beyond just scientific teaching and a communal ski trip."**
> — Ulrike Beisiegel

This year's winners of *Nature*'s Awards for Mentoring in Science (from top): Heinrich Betz, Peer Bork and Klaas Bergmann.

MOVERS

Julio Frenk, dean, Harvard School of Public Health, Boston, Massachusetts

2007-08: Senior fellow, Global Health Program, Bill & Melinda Gates Foundation, Seattle, Washington; President, Carso Health Institute, Mexico City, Mexico
2000-06: Minister of health, Mexico City, Mexico
1998-2000: Executive director, evidence and information policy, WHO, Geneva, Switzerland

Medical student Julio Frenk was inspired to address health policy in his native Mexico by a visit to the impoverished southern state of Chiapas. After a career immersed in public health, Frenk now has a global reach. Colleagues say the newly appointed dean of the Harvard School of Public Health has a superlative track record of creating health policy based on scientific evidence.

After receiving his medical degree from the National Autonomous University of Mexico, Frenk pursued a public-health career at the University of Michigan. He obtained master's degrees in public health and in sociology, but it was a PhD in medical-care organization and sociology that gave him insight into crafting health-care policy.

While working on his degrees, Frenk wrote articles critical of Mexico's medical care that caught the eye of the new health minister, Guillermo Soberón, who was eager to improve the country's epidemiological capability. Frenk developed a proposal for a Center for Public Health Research and became the founding director first of that, and then of the National Institute of Public Health.

"This opportunity to combine excellence with relevance was the beginning of my career," he says. "This is a great example of how a world-class university can help build a developing country's capacity — without creating dependence."

Harvey Fineberg, then dean of the Harvard School of Public Health and one-time adviser to Mexico's nascent institute, says that Frenk seemed destined to excel in public health and medicine because of his broad grasp of the issues. "He has converted evidence into practice with a strength and vision seldom combined," says Fineberg, now president of the US Institute of Medicine.

Later, a report from Frenk on the Mexican health system and recommendations necessary for health-care reform caught the attention of the World Health Organization's then director-general, Gro Harlem Brundtland. She hired Frenk to do similar work at a global level as executive director of the organization's evidence and information-policy section. In 2000, Frenk became Mexico's minister of health, enacting reforms based on his analyses.

At Harvard, Frenk plans to explore what public health should look like in the twenty-first century. "A citizen of Mexico will bring a fresh perspective and send a very important signal that Harvard is serious about its global reach," says Fineberg.
Virginia Gewin

NETWORKS & SUPPORT

Getting and gaining from interviews

Securing a faculty position requires hard work and a little luck. Based on our recent job-hunting experiences, we offer these tips to help increase the chance of getting an interview and receiving a job offer.

Tailor your covering letters to match the job description, and to convey knowledge of the department to which you are applying. When we were applying, this helped us get interviews in areas peripherally related to our disciplines. Note also that three or four first-authored publications in leading journals from PhD and postdoc work can be sufficient to secure an interview.

Fill half your research statement with achievements, and the other half with a clear description of realistic goals. Get these statements critically reviewed by postdocs and faculty members from diverse disciplines.

Remember that reference letters will not always be glowing. When in doubt, ask if the referee is willing to write you a strong letter. This can avoid wasting six months without an interview. Give referees sufficient time, send gentle reminders and, as a back-up, ask one or two others if they are willing to write a letter for you.

Networking is useful: meeting seminar speakers, giving presentations at meetings and e-mailing colleagues can all help to get you an interview.

The skills for securing an interview are different from those needed in the interview itself. Giving a great seminar is key. Practise it in front of diverse, critical colleagues to help you identify potential points of confusion for the audience and weaknesses in your research plans. Also make sure you work out what the interviewers will expect to gain from the seminar and the likely composition of the audience you will have.

Make sure that the plans for your future research are clear. The most common questions we were asked related to specific goals, first grants and funding sources, projects for students at all levels and major equipment needs.

Interviewers ask themselves "Would this person be a good colleague?" So it is crucial to be able to hold a conversation, while showing interest in and knowledge of others' work.

Ask questions. Ours focused on collegiality, teaching and tenure requirements, student quality, departmental infrastructure, gender equality and parental policies. And, most of all, remember to show enthusiasm. ■
Siobhán Brady and Marc Johnson are biology postdocs at Duke University in Durham, North Carolina.

POSTDOC JOURNAL

Lessons from Formula One

"It was a great fight and I don't think there was anything wrong," announced racer Lewis Hamilton, who was accused of cutting a corner at the Belgian Formula One Grand Prix. As I watched, I mused about corner-cutting in science, and whether such practices are justified or even necessary in order to succeed.

When data are presented, the reader or listener assumes they are robustly reproducible. One trusts that quantitative results are based on an adequate number of experimental replicates and reproducible results, and that the design includes appropriate controls. Are such assumptions necessarily valid? Much may be left unsaid, especially in a culture in which it is important to save face.

A student from another lab once sought my advice on alternative experimental approaches, claiming that her original one had failed. I later discovered that she had attempted the experiment only once, and without proper controls. Even in the collegial atmosphere of lab meetings, there is pressure to look good in front of both peers and supervisor. The emphasis on positive data is quite strong. Negative data, technical problems and methodological shortcomings may be overlooked.

Hamilton was penalized for his alleged corner-cutting. But short cuts in the lab may never be detected — even though they could matter a great deal. ■
Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Invites applications for a
Science Officer
Peer Review
in the Chief Executive's unit

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance European research and explore new directions for research at the European level.
Established in 1974 as an independent non-governmental organisation, ESF currently serves 77 Member Organisations (Research Funding Agencies, Research Performing Organisations and Academies) across 30 countries.

Mission of the Position
The mission of the position is developing, promoting and delivering, within the strategy set by ESF's Governance, ESF's expanding role as a European-level provider of high quality peer review for its member organisations and more widely. The Science Officer will report to the Head of the Chief Executive's Unit, and will work in close collaboration with the Senior Science Officer responsible for Peer Review and Quality Assurance.

Position Responsibilities
This position will involve:
• Acting as "development manager" for the expansion and professionalisation of ESF's external peer review activities;
• Advising the management on the priorities/planning of peer review requests from MO's or others;
• Supporting high quality implementation of ongoing peer review activities in ESF scientific instruments, and other approved ESF activities;
• Implementing scientific quality control, harmonisation and standards, and thereby guaranteeing sound, efficient, and robust peer review processes and evaluation of ongoing and completed activities;
• Liaising with ESF Member Organisations, COST and external research supporting bodies;
• Involving and integrating the ESF scientific Units, their committees, and where appropriate their research communities, in the development and delivery of ESF's Peer Review Strategy;
• Publicising and informing ESF's community (presentations, writing material for publications and the web) and liaising with the ESF Communications Unit;
• Management of specific activities and their budgets Supporting the Head of Unit and the Senior Science Officer responsible for Peer Review in the delivery of the Units mission.

The post is based at ESF in Strasbourg, France.
Profile and working conditions are described in the complete position announcement available at www.esf.org/vacancy-ceprso
Please send your application by **24 November** 2008 to **jobs@esf.org** quoting the following reference identifier **CEPR–SO**.
Interviews will be held in Strasbourg **beginning of December 2008.**

Further details at **www.esf.org**

W172695R

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Invites applications for a
Science Officer Member
Organisation Fora
in the Chief Executive's Unit

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance European research and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, ESF currently serves 77 Member Organisations (Research Funding Agencies, Research Performing Organisations and Academies) across 30 countries.

Mission of the Position
MO Fora are a strategic instrument of ESF which provide a mechanism for ESF Member Organisations, and other important key stakeholders, to discuss and coordinate strategic and operation matters of joint interest. The Fora provide the opportunity to increase coordination of national activities, to develop transnational strategies and policies, to promote synergy and to exchange best practice. MO Fora are a crucial element in the work of the ESF and its Member Organisations in developing the future of the European Research Area (ERA).
See **http://www.esf.org/activities/mo-fora.html** for further details.
The primary mission of the position is to take responsibility, under the direction of the Head of the Chief Executive's Unit, for supporting existing MO Fora, for instance in Research Integrity and Research Infrastructures, so that they succeed in delivery their specific objectives. The position will also take an important role in the identification development and delivery of new topics.

Position Responsibilities
This position will involve:
• Implementing and broadening the ESF MO Fora instrument;
• Facilitating the progress of individual MO Fora by initiating and coordinating meetings and activities, writing reports etc;
• Liaising closely with colleagues in ESF Member Organisations, and other research supporting bodies;
• Working within the Chief Executive's Unit on related policy and strategy activities in relation to Member Organisations, and other ERA stakeholders, for example on research infrastructures issues;
• Publicising and informing ESF's community (presentations, writing material for publications and the web) and liaising with the ESF Communications Unit;
• Management of specific activities and their budgets;
• Supporting the Head of Unit in the delivery of the Units mission.

The post is based at ESF in Strasbourg, France.
Profile and working conditions are described in the complete position announcement available at www.esf.org/vacancy-cemofso
Please send your application by **15 December** 2008 to **jobs@esf.org** quoting the following reference identifier **CEMOF–SO**.
Interviews will be held in Strasbourg on **January 2008**.

Further details at **www.esf.org**

W172693R

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Invites applications
for the position of
Science Officer
to the Chief Executive

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance European research and explore new directions for research at the European level.
Established in 1974 as an independent non-governmental organisation, the ESF currently serves 77 Member Organisations including Research Funding Agencies, Research Performing Organisations and Academies, across 30 countries.

Mission
The post holder will work in direct support of the Chief Executive, notably on interaction with ESF Member Organisations, the European Commission and other ERA stakeholders.

Position Responsibilities
This position will involve:
• Supporting the Chief Executive in interfacing with external Organisations, particularly Member Organisations and other major ERA stakeholders;
• Developing personal networks of contacts within the Member Organisations and other ERA stakeholders, to create awareness of issues and to identify or create opportunities to progress the ESF's agenda and influence policy development;
• Accompanying the Chief Executive to business meetings and ensuring that any actions arising are progressed;
• Assisting in the writing of papers for the governing bodies of the ESF;
• Preparing briefing, talks and presentations for the Chief Executive;
• Supporting the Chief Executive as secretary to the meetings of the Science Advisory Board, of the Chairs of Committees of the ESF and other meetings as required;
• Liaising with other ESF Units and external bodies on science policy issues;
• Supporting other strategy and policy tasks as required;
• The post is based in the Chief Executive's Unit.

The post is based at ESF in Strasbourg, France.
Profile and working conditions are described in the complete position announcement available at www.esf.org/vacancy-soce
Please send your application by **1 December** 2008 to **jobs@esf.org** quoting the following reference identifier **SO–CE**.
Interviews will be held in Strasbourg **beginning of December 2008**.

Further details at **www.esf.org**

W172706R

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Invites applications for a
Science Officer
Forward Looks
in the Chief Executive's unit

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance European research and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, ESF currently serves 77 Member Organisations (Research Funding Agencies, Research Performing Organisations and Academies) across 30 countries.

Mission of the Position
ESF's Forward Looks are the principal strategy instrument of the ESF, seeking to identify key future directions and goals in research and research strategy at the European level and the means to achieve those goals. Recognising the high priority of this activity, the ESF's Member Organisations have recently approved a significant increase in investment to ensure and enhance the impact of the Forward Look instrument. This position is a direct result of this new investment.
The position's mission is to further develop, promote and ensure delivery, within the strategy set by ESF's Governance, of ESF's Forward Look instrument. The Science Officer will report to the Head of the Chief Executive's Unit, and will work in close collaboration with the Director Science and Strategy, and with ESF's science units.

Position Responsibilities
This position will involve:
• Developing and managing the Forward Look instrument, including refining the protocols for initiating, selecting and implementing Forward Look projects;
• Managing calls and suggestions for Forward Look proposals, and the subsequent assessment and selection procedures to ESF's Quality Assurance standards;
• Coordinating the implementation of awarded projects and the dissemination of results;
• Working closely with ESF's scientific units, their committees, and where appropriate research communities and external consultants, in the development and delivery of ESF's Forward Look Strategy;
• Liaising with ESF Member Organisations, COST and external research supporting bodies;
• Publicity and information for ESF's community on the Forward Look instrument (presentations, writing material for publications and the web) in liaison with the ESF Communications Unit;
• Management of specific activities and their budgets;
• Supporting the Head of Unit in the delivery of the Units mission.

The post is based at ESF in Strasbourg, France.
Profile and working conditions are described in the complete position announcement available at www.esf.org/vacancy-ceflso
Please send your application by **24 November** 2008 to **jobs@esf.org** quoting the following reference identifier **CE–FLSO**.
Interviews will be held in Strasbourg **early December 2008**.

Further details at www.esf.org

W172692R

Faculty Position in BioMicroengineering
at the Ecole polytechnique fédérale de Lausanne (EPFL)

The Institutes of Bioengineering and Microengineering at EPFL are seeking a **tenure track assistant professor** in the field of **biomicro/nanosystems.** Exceptionally qualified candidates may also be considered at a more senior level. Bioengineering at EPFL is well integrated between the Schools of Engineering and Life Sciences. In this search, an appointment is sought within the School of Engineering, in which Microengineering is also situated. The open faculty position is offered in an environment of both theoretical and experimental research, rich for the development of novel basic technologies as well as for seeking deeper understanding of integrative (patho)physiological mechanisms and developing novel technological and biotherapeutic approaches at the levels of genes, biomolecules, cells and tissues.

The Institutes seek to grow at the interface of engineering with biology in the domain of nano- and microtechnologies and integrated systems. We particularly encourage candidates with strong expertise in the areas of bio-MEMS/NEMS, sensing and actuation for in vitro use, for example in basic biological investigation, systems biology, and diagnostics, or in vivo use, for example in diagnostic and sensing systems, integrated systems, and drug delivery. The interface with robotic, surgical and imaging systems is of specific interest. EPFL has strong research facilities, in particular in micro/nanofabrication, imaging, and cytometry.

Successful candidates are expected to initiate independent, creative research programs and participate in undergraduate and graduate teaching. We offer internationally competitive salaries, start-up resources and benefits.

Applications should include a curriculum vitae with a list of publications, a concise statement of research and teaching interests, and the names and addresses (including e-mail) of at least five referees. Applications should be uploaded to **http://biomems-rec.epfl.ch.** The deadline for applications is **15 January 2009.**

Enquiries may be addressed to:
Prof. Jeffrey A. Hubbell,
e-mail: **biomems-rec@epfl.ch**

For additional information on EPFL, School of Engineering, Institute of Bioengineering, and Institute of Microengineering, please consult the web sites:
**http://www.epfl.ch, http://sti.epfl.ch,
http://ibi.epfl.ch, http://imt.epfl.ch**

EPFL aims to increase the presence of women amongst its faculty, and qualified female candidates are strongly encouraged to apply.

W172174R

Naturejobs editorial overview

Prospects
A quick take on how headlines affect science jobs

•

Special Reports
Issues and alternatives for the research professional

•

Careers and Recruitment
Global opportunities in different disciplines

•

Spotlight/Regions
A tour of scientific hubs

•

Career View
The voice of organizations across the globe

www. naturejobs .com

naturejobs
making science work

Millions of people need you to do a brilliant job.

We want you to say yes.

Creative people with know-how and ideas in R&D are the driving force of a new generation in biopharma. We have a clear vision of tomorrow´s activities: shifting the focus to the central nervous system and immunology disorders - an innovative science calling for the best minds from around the world. So we´re looking far beyond the borders to assemble a large number of new professionals. Together we can significantly improve the quality of life for patients with serious diseases. So they can enjoy a normal life again.

You can play a crucial role in this effort if you´re determined to do intensive research as part of the collaborative development of new therapies. Certainly if you have a Ph.D. plus three to five years of practical experience and specialisation in biology, pharmacology, pharmacokinetics or chemistry. At UCB, you don´t work alone, but in a team that succeeds together. An ideal climate for encouraging dialogue among specialists, but also for listening to your ideas. This makes the contribution of each colleague especially valuable. By combining our strengths, we can have an impact on the lives of millions of people, day after day.

Want to be part of the new generation? Let yourself be inspired by www.ucb-group.com

www.ucb-group.com

McCANN | PEOPLE

ucb

The next generation biopharma leader

W172762R

OPPORTUNITY AT THE UNIVERSITY OF GENEVA

THE FACULTY OF MEDICINE of the UNIVERSITY of GENEVA
is seeking applications for a position of:

FULL OR ASSOCIATE PROFESSOR OF INTEGRATED METABOLISM

in the Department of cell physiology and metabolism

The host Institution has a long standing tradition of excellence in diabetes research and islet biology. The present call aims at reinforcing strengths in integrated metabolism.
Candidates should have a recognized expertise in the field of integrated metabolism related to diabetes and/or obesity research.

This position involves responsibilities for teaching metabolism at the graduate and post graduate levels.

Applicants should be able to direct a competitive research program in a particular area of the field, coordinate research projects in collaboration with other medical specialities, and assume pertinent administrative tasks.

A Doctorate of Medicine or Biology (MD or PhD) or equivalent degree is required; knowledge of the French language would be an asset.

The starting date for the position is April 1st 2010, or according to agreement.

Information concerning applications and job description are available from Stephane.jouve@unige.ch - Tel. +41 22 379 50 05 – Fax : +41 22 379 50 02

Applications must be sent before the January 31st, 2009 to:
Prof. Jean-Louis CARPENTIER, Dean
Faculty of Medicine, University of Geneva
Décanat CMU
1 rue Michel-Servet,
CH-1211 Genève 4
Switzerland

Women are encouraged to apply

UNIVERSITÉ DE GENÈVE
FACULTÉ DE MÉDECINE

W171858R

Fonds National de la Recherche Luxembourg

ATTRACT RESEARCH GRANT
Opportunities for Outstanding Young Researchers in Luxembourg

What is it? A grant programme by the National Research Fund Luxembourg (FNR) which offers outstanding international researchers the opportunity to set up an independent research team within a public-sector research institution in Luxembourg.

How does it work? Candidates jointly submit a project proposal together with a Luxembourg public-sector research institution. The FNR chooses up to two candidates per call. Funding is allocated for five years and projects may obtain up to EUR 1,000,000 as a contribution from the FNR.

Who may apply? Candidates must be excellent and have gained a minimum of two and a maximum of eight years' professional experience since successful completion of their doctoral studies.

Interested? A list of Luxembourg research institutions and priority research domains can be found on our web site **www.fnr.lu**.

Call Deadlines:
2 February 2009: Submission of summary proposals
15 April 2009: Submission of full proposals

For further information please contact:
Mr Frank Glod, Programme Manager
Phone: +352 26 19 25 33
Email: frank.glod@fnr.lu

W171927R

Do you have science
career questions?
Join-in the discussion with the new
Naturejobs group on Nature Networks
network.nature.com/group/naturejobs

naturejobs
making science work

Change your environment. Find jobs where you'll make a difference

naturejobs

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Faculty Position in Materials Science
at the Ecole polytechnique fédérale de Lausanne (EPFL)

The School of Engineering of EPFL invites applications for a faculty position in its Institute of Materials to begin during the calendar year 2009. The opening is for a position at the **tenure-track assistant professor** level. We seek outstanding individuals who will develop and drive a research program at the forefront of the discipline, as well as contribute to curriculum development and teaching in the Bachelor, Master and Doctoral academic programs.

Top-level applicants in Materials Science and Engineering having expertise in the general area of **interfaces and multi-materials** will be considered. Topics of interest in this domain include, but are not limited to: designed multiphase materials for structural, functional, bio-related, energy-related, sustainability or recycling applications; methodologies, both theoretical and experimental, enabling the design of interfaces having tailored functionality; novel approaches towards the processing of nanoscale and hybrid materials.

Significant start-up resources and state-of-the-art research infrastructure will be available. Salaries and benefits are internationally competitive.

Applications should be submitted via the web site **http://imx-search08.epfl.ch** and should include the following documents in PDF format: curriculum vitae, publication list, brief statement of research and teaching interests, names and addresses (including e-mail) of 6 references.

The deadline for applications is **15 January 2009**.

Enquiries may be addressed to:
Prof. Andreas Mortensen
E-mail: **hiring.imx@epfl.ch**

For additional information on EPFL, please consult the web sites **http://www.epfl.ch**, **http://sti.epfl.ch** and **http://imx.epfl.ch**.

EPFL aims to increase the presence of women amongst its faculty, and qualified female candidates are strongly encouraged to apply.

W172171R

ENP
École des Neurosciences Paris Île-de-France

Neuroscience

The Paris School of Neuroscience (ENP) is a network of outstanding laboratories in the Paris area, within major universities and research institutes. The ENP aims at facilitating graduate training and promoting cutting edge research. The scope of the ENP covers all areas of Neuroscience and associated methodologies, from fundamental to clinical and from molecular to cognitive.

The ENP offers:
- Doctoral training: Applications in September and March,
- Positions for outstanding senior and junior group leaders and post-doctoral fellows,
- Summer schools and thematic international meetings.

Founding institutions:
CEA, CNRS, Inserm, Université Pierre & Marie Curie (Paris 6), Université Paris-Sud 11.

Host institutes:
Collège de France, École Normale Supérieure, Institut du Cerveau et de la Mœlle, Alfred Fessard Institute, Fer à Moulin Institute, Pasteur Institute, Salpêtrière Hospital, Neurospin, Campuses of Universities Paris 5, 6, 11, and other institutions.

For further information:
Web: http://www.paris-neuro science.fr/en/enp/index.php

W171549R

ESRF
European Sychrotron Radiation Facility

We Highlight Science
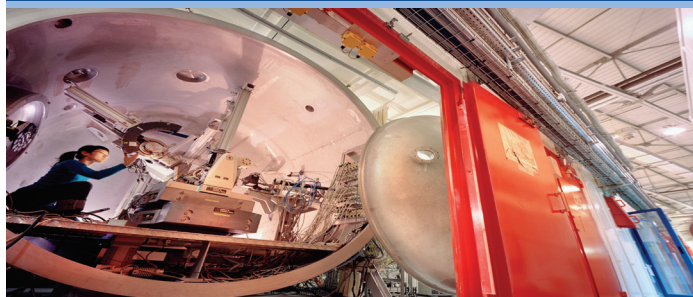
Physics and Chemistry, Life Sciences and Medicine, Earth and Environmental Sciences, Surface and Materials Sciences.

The European Synchrotron Radiation Facility (ESRF) is Europe's most powerful light source. The ESRF offers you an exciting opportunity to work with international teams using synchrotron light in Grenoble, in the heart of the French Alps.

Have a look at our vacancies at
www.esrf.eu/jobs

Contact us at **recruitment@esrf.eu**

Scientists - Post doctoral fellows - PhDstudents - Engineers - Technicians - Administrative staff

European Sychrotron Radiation Facility
ESRF, BP 220, F-38043 Grenoble Cedex 9, FRANCE,
Tel.+33 476 88 20 00 www.esrf.eu

W172400R

Ateliers de formation Inserm
101 rue de Tolbiac
75654 Paris Cedex 13 France

Tel: 33 (0) 144.23.62.04
Fax: 33 (0) 144.23.62.93
ateliers@inserm.fr

Instituts thématiques

Inserm

French Institute
of Health and Medical Research

Ateliers de formation 2009

192 Human pluripotential stem cells

Phase I Critical assessment: March 8-10, 2009 Saint-Raphael
Organizers: Claire Rougeulle (UMR Epigénétique et Destin Cellulaire, Institut Pasteur, Paris), Marc Lalande (University of Connecticut Stem Cell Institute, Farmington, USA).
Tentative list of speakers: Peter Andrews (Sheffield, UK), Lyle Armstrong (Newcastle, UK), Véronique Azuara (London, UK), Annelise Bennaceur (Villejuif, France), Oliver Brüstle (Bonn, Germany), Carine Camby (Paris, France), Chad Cowan (Boston, USA), John de Vos (Montpellier, France), Kevin Eggan (Boston, USA), Alison Murdoch (Newcastle, UK), Marc Peschanski (Evry, France), Benjamin Reubinoff (Jerusalem, Israel), Ludovic Vallier (Cambridge, UK), Ren-He Xu (Farmington, USA), Franck Yates (Villejuif, France), Lorraine Young (Nottingham, UK).

Phase II November 9-13, 2009 University of Connecticut, Farmington, USA
Registration deadline: December 19, 2008

193 Polymorphism and genomic rearrangements: analysis of CGH and SNP array data, and deep sequencing data

Phase I Critical assessment: April 15-17, 2009 Saint-Raphael
Organizers: Emmanuel Barillot (Institut Curie, Paris), Yves Moreau (Leuven University, Belgium)
Tentative list of speakers: Alain Aurias (Paris, France), Olivier Delattre (Paris, France), Richard Durbin (Cambridge, UK), Janet Fridlyand (San Francisco, USA), Philippe Hupé (Paris, France), Olli Kallioniemii (Turku, Finland), Björn Menten (Ghent, Belgium), Yves Moreau (Leuven, Belgium), H.-Hilger Ropers (Munchen, Germany), Steven Scherer (Toronto, Canada), Simon Tavaré (Cambridge, UK), Joris Veltman (Nijmegen, The Netherlands), Joris Vermeesch (Leuven, Belgium), Martin Vingron (Munchen, Germany), Bauke Ylstra (Amsterdam, The Netherlands).

Phase II Technical Workshop: June 9-10, 2009 Paris
Registration deadline: February 6, 2009

194 Tissue Engineering: study of the interfaces cell/tissue/material

Phase I Critical assessment: May 27-29, 2009 Saint-Raphael
Organizers: Joëlle Amédée (Inserm U577, Bordeaux), Jérôme Guicheux (Inserm U791, Nantes), Didier Letourneur (Inserm U698, Paris)
Tentative list of speakers: Karine Anselme (Mulhouse, France), Mario Barbosa (Porto, Portugal), Odile Damour (Lyon, France), Nicolas L'Heureux (Novato, USA), Laurent Laganier (Mions, France), Patrice Laquerrière (Reims, France), Philippe Lavalle (Strasbourg, France), Didier Mainard (Nancy, France), Ivan Martin (Basel, Switzerland), Josep A. Planell (Barcelona, Spain), Luc Sensebe (Tours, France), Clemens van Blitterswijk (Enschede, The Netherlands), Pierre Weiss (Nantes, France).

Phase II Technical workshop: September 2-4, 2009 Nantes
Registration deadline: March 20, 2009

195 Novel imaging techniques for biology: super-resolution and super-localization

Phase I Critical assessment: June 3-5, 2009 Saint-Raphael
Organizers: Benoît Dubertret (ESPCI, Paris), Olivier Haeberle (Université de Haute Alsace, Mulhouse), Vincent Loriette (ESPCI, Paris)
Tentative list of speakers: Joerg Bewersdorf (The Jackson Laboratory, USA), Laurent Cognet (Bordeaux, France), Rainer Heintzmann (King's College, UK), Lars Kastrup (Goettingen, Germany), Jérôme Mertz (Boston University, USA), Mark Neil (Imperial College, UK), Raimund Ober (University of Texas, USA), Gleb Shtengel (Howard Hughes Medical Institute, USA), Jean-Baptiste Sibarita (Paris, France), Jean-Luc Vonesch (Strasbourg, France), Tony Wilson (Oxford, UK).

Phase II Technical workshop : July 2009 Paris/Bordeaux
Registration deadline: March 20, 2009

196 Ubiquitin, ubiquitin-like proteins, and Proteasomes: functions and dysfunctions

Phase I Critical assessment: June 10-12, 2009 Saint-Raphael
Organizers: Olivier Coux (Montpellier, France), Catherine Dargemont (Institut Jacques Monod, Paris)

Tentative list of speakers: Olivier Coux (Montpellier, France), Catherine Dargemont (Institut Jacques Monod, Paris), Mickaël Glickman (Haïfa, Israel), Fred Golberg (Boston, USA), Ron Hay (Dundee, UK), Jon Huibrebregtse (Austin, USA), Alain Israel (Paris, France), Stefan Jentsch (Munnich, Germany), Frauke Melchior (Göttingen, Germany), Martin Scheffner (Constance, Germany), Thomas Sommer (Berlin, Germany), Keiji Tanaka (Tokyo, Japan), William Tansey (Cold Spring Harbor, USA), Rosine Tsapis (Paris, France)

Phase II Technical Workshop: October 2009 Paris/Montpellier
Registration deadline: March 27, 2009

197 Metabolic and structural exploration of mitochondria in pathology and therapeutic perspectives

Phase I Critical assessment: September 16-18, 2009 Saint-Raphael
Organizers: Jean-Pierre Mazat (Inserm U688, Bordeaux), Vincent Procaccio (University of California, CA/USA), Pascal Reynier (Inserm U694, Angers)
Tentative list of speakers: Roderick Capaldi (Eugene, OR/USA), Arnaud Chevrollier (Angers, France), Jean-Paul di Rago (Bordeaux, France), Chittibabu Guda (Rensselaer NY/USA), Marcia Haigis (Boston, MA/USA), Guy Lenaers (Montpellier, France), Anne Lombes (Paris, France), Carmen Mannella (Albany, NY/USA), Jean-Pierre Mazat (Bordeaux, France), Arnold Munnich (Paris, France), Vincent Procaccio (Irvine, CA/USA), Thierry Rabilloud (Grenoble, France), Manuel Rojo (Bordeaux, France), Pierre Rustin (Paris, France), Douglas Wallace (Irvine, CA/USA).

Phase II Technical workshop: September 21-23, 2009 Bordeaux
Registration deadline: June 26, 2009

198 Recent study designs in epidemiology

Phase I Critical assessment: September 30- October 2nd, 2009 Saint-Raphael
Organizers: Nadine Andrieu (Inserm U900, Paris), Michel Chavance (Inserm U780, Villejuif), Pascal Wild (INRS, Nancy).
Tentative list of speakers: Nadine Andrieu (Paris, France), Norman Breslow (Seattle, USA), Michel Chavance (Villejuif, France), Patrick Farrington (Milton Keynes, USA), Bryan Langholz (Los Angeles, USA), Thomas Lumley (Seattle, USA), Helena Marti-Soler (Villejuif, France), Walter Schill (Brême, Germany), Pascal Wild (Nancy, France).
Registration deadline: July 17, 2009

199 Human memory and its impairment: multidisciplinary approach

Phase I Critical assessment: October 7-9, 2009 Saint-Raphael
Organizers: Béatrice Desgranges (Inserm U923, Caen), Francis Eustache (Inserm U923, Caen), Bernard Laurent (Hôpital de Bellevue, Saint-Etienne).
Tentative list of speakers: Hélène Amiéva (Bordeaux, France), Sylvie Belleville (Montréal, Canada), Gaël Chételat (Caen/France, Melbourne/Australia), Julien Doyon (Montréal, Canada), Francis Eustache (Caen, France), Bernard Laurent (Saint-Etienne, France), Stéphane Léhéricy (Paris, France), Pascale Piolino (Caen/Paris, France), Michel Poncet (Marseille, France), Bruno Poucet (Marseille, France), Catherine Thomas-Antérion (Saint-Etienne, France), Julie Snowden ((Manchester, UK), Martial van der Linden (Geneva, Switzerland).
Registration deadline: July 7, 2009

200 Functional organization of genomes in the nucleus: from molecular to in vivo approaches

Phase I Critical assessment: October 19-21, 2009 Saint-Raphael
Organizers: Frédéric Bantignies (IGH, Montpellier), Angela Taddei (Institut Curie, Paris).
Conférenciers pressentis Tentative list of speakers: Geneviève Almouzni (Paris, France), Giacomo Cavalli (Montpellier, France), Xavier Darzacq (Paris, France), Job Dekker (Worcester, MA/USA), Wouter de Laat (Rotterdam, The Netherlands), Christophe Escudé (Paris, France), Thierry Forné(Montpellier, France), Susan Gasser (Basel, Switzerland), Edith Heard (Paris, France), Terumi Kohwi-Shigematsu (Berkeley, CA/USA), Ulrich Laemmli (Geneva, Switzerland), Rolf Ohlson (Uppsala, Sweden), Yijun Ruan (Singapore, Singapore), Remi Terranova (Basel, Switzerland), Bas van Steensel (Amsterdam, The Netherlands).
Registration deadline: July 10, 2009

W171019A

National Institute of Mental Health
Division of Intramural Research Programs
Mood and Anxiety Disorders Program
Tenured Clinical Investigator

The Mood and Anxiety Disorders Program, Division of Intramural Research Programs (DIRP), National Institute of Mental Health (NIMH), National Institutes of Health (NIH), seeks a highly accomplished Tenured Clinical Investigator conducting clinical trials in mood disorders to head an active, ongoing program in this area which will have access to an inpatient ward with 10-12 beds in the NIH Clinical Center, the nation's largest hospital devoted entirely to clinical research. The position comes with a budget and staff. The strong scientific environment and outstanding resources at NIMH make this a unique opportunity for a high-achieving scientist. The position also offers unparalleled opportunities for interdisciplinary collaboration with scientists throughout the NIH. The successful candidate will be expected to strengthen the current program.

Applicants should have: 1) an M.D. degree, and be board certified in Psychiatry; 2) experience (including protocol development, implementation, drug development, single-site studies, publications, etc.) as an investigator in translational clinical research. Such experience would include directly applying relevant preclinical findings into randomized placebo-controlled proof-of-concept trials with novel compounds in treatment-resistant depression and mania; 3) experience in biomarker studies, applying the techniques of neuropsycho-pharmacology, electrophysiology, neuropsychology, and neuroimaging to research involved in the development of novel therapeutics for mood disorders; 4) national/international recognition for experimental therapeutic studies in mood disorders; and 5) experience administering a therapeutic development research program.

The ideal candidate must have a record of high scientific achievement as an independent investigator. The candidate should also have substantial clinical achievements, as well as leadership experience, including oversight of team members. Finally, applicants should have experience directly administering a research program. These achievements should be nationally and internationally recognized.

Salary is commensurate with experience and accomplishments, and a full Civil Service package of benefits (including retirement, health, life, and long-term care insurance, as well as a Thrift Savings Plan, etc.) is available. NIMH is a major research component of the National Institutes of Health and the Department of Health and Human Services, which have nationwide responsibility for improving the health and well-being of all Americans. Interested applicants should send curriculum vitae, bibliography, statement of research interests, accomplishments, and goals, together with six letters of reference to: **Dr. Judith Rapoport, Chair, Search Committee for Experimental Therapeutics, NIMH, NIH, Bldg. 10, Rm. 4N-222, 9000 Rockville Pike, Bethesda, MD 20892; or e-mail to: steyerm@mail.nih.gov** Review of applications will begin **January 5th, 2009**, but applications will continue to be accepted and considered until the position is filled.

National Institute of Mental Health
Deputy Director
Office of the Director

The National Institute of Mental Health, a major research component of the National Institutes of Health (NIH) and the Department of Health and Human Services (DHHS), is seeking exceptional candidates for the position of Deputy Director, Office of the Director (OD). The Deputy Director serves as the second-in-command for the Institute. Working closely with the Director, the Deputy Director assists in the scientific and administrative management of an organization with a budget of $1.4 billion and a staff of approximately 1,300. **(http://www.nimh.nih.gov/index.shtml)**

The Deputy Director is primarily responsible for implementation of the Institute's Strategic Plan and management of the daily operations of the Institute. In this sense, the Deputy Director is the chief internal champion and guarantor for the intellectual and administrative environment at the Institute. The Deputy Director is aided in this effort by the Directors of the Institute's divisions and offices. The Deputy Director also serves as an ambassador and spokesperson for the Institute.

Applicants must have a Ph.D., M.D., or equivalent degree in the biomedical sciences, with broad senior-level research experience and experience in direct administration of a research program. Applicants should be known and respected within their profession, both nationally and internationally, as distinguished individuals of outstanding scientific competence and administrative capability. Salary is commensurate with experience and accomplishments. Experience with NIH administrative policies, procedures, and operations is highly desirable but not essential.

Interested candidates should send a letter of interest, including a brief description of research and administrative experience, a curriculum vitae and bibliography, and the names of at least three references to: **Chair, NIMH Deputy Director Search Committee at NIMHsearch@mail.nih.gov or at 6001 Executive Blvd, Room 8235, MSC 9669 Bethesda, MD 20892-9669 (for express or courier delivery use Rockville, MD 20852)**. Review of applications will begin on **November 3, 2008**, but applications will continue to be accepted and considered until the position is filled. For questions contact **Dr. Thomas Insel, Director, NIMH at tinsel@mail.nih.go**v

---

The Neuro-Oncology Branch, a trans-institute program of the National Cancer Institute and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health, is recruiting a Staff Scientist to work in the area of cancer genomics and/or experimental therapeutics. The successful candidate should have a M.D. or Ph.D. degree, and at least 3 years of post-doctoral training in molecular and cellular biology. Laboratory projects include the genetic study of primary glial neoplasms and neural and tumor stem cells for the purpose of identifying novel anti-tumor targets, and the development of high throughput screens for small molecule inhibitors of tumor/ stem cell-associated signal transduction pathways. Experience in cancer and/or stem cell biology would be desirable.

STAFF SCIENTIST

Please send curriculum vitae, statement of research interests and two letters of reference to:
Karen B. Abraham
Administrator, Neuro-Oncology Branch, National Cancer Institute
MSC 8200, Room 225
9030 Old Georgetown Road
Bethesda, MD 20892-8200
abrahamka@mail.nih.gov

Investigator Recruitment in Cancer Genetics
National Human Genome Research Institute

The Cancer Genetics Branch (CGB) of the National Human Genome Research Institute (NHGRI) is seeking to recruit an outstanding tenure-track investigator to pursue innovative, independent research in cancer genetics. General areas of interest include, but are not limited to:

- **Cancer Gene Discovery**
- **Comparative Cancer Genomics**
- **Genetic Epidemiology**
- **Molecular Profiling of Tumors**
- **Functional Genomics of Cancer**
- **Genome Instability in Cancer**
- **Markers for Early Detection**
- **Genetics of Tumor Progression**

The successful candidate will be able to take advantage of interactions with a highly collegial group of scientists within NHGRI and on the NIH campus as a whole. In addition, they will have access to NHGRI's outstanding core laboratories.

Candidates must have a Ph.D., M.D., or equivalent degree, as well as comprehensive, advanced training and a record of accomplishment in one of the targeted areas. This position includes generous start-up funds, an ongoing commitment of research space, laboratory resources, and positions for personnel and trainees.

Interested applicants should submit a *curriculum vitae*, a three-page description of their proposed research, and three letters of recommendation through our online application system, at http://research.nhgri.nih.gov/apply.

Applications will be reviewed starting November 21, 2008 and will be accepted until the position is filled.

For more information on CGB and NHGRI's Intramural Program, please see http://genome.gov/DIR. Specific questions regarding the recruitment may be directed to Dr. Joan Bailey-Wilson (Search Chair) at jebw@nhgri.nih.gov or by fax at 410-550-7513. Questions may also be directed to Dr. Elaine Ostrander, Chief, Cancer Genetics Branch, at eostrand@mail.nih.gov or by fax at 301-480-0472.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE   Division of Intramural Research
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

Investigator Recruitment in Genetic Disease Research
National Human Genome Research Institute

The Genetic Disease Research Branch (GDRB) of the National Human Genome Research Institute (NHGRI) provides unparalleled opportunities for young investigators to develop world-class research programs in genetics and genomics. The Branch is pleased to announce that it is seeking to recruit a new tenure-track investigator to pursue innovative, independent research as part of this group of highly interactive and supportive investigators.

Current GDRB faculty members use a variety of approaches to study the regulation and function of genes involved in normal and abnormal development, focusing on diseases in both humans and model systems. We are seeking to recruit an individual whose research interests and approaches complement those already found within the Branch. Specifically, the ideal candidate will have an interest in developing a research program that integrates:

- **Clinical or translational research**
- **Molecular and genomic approaches aimed at understanding the mechanisms of normal development and disease**
- **Basic genetic or genomic research**

The Branch strongly supports interdisciplinary research, with NHGRI faculty providing mentoring and guidance to individuals interested in developing research programs involving basic, clinical, and translational approaches.

The successful candidate will be able to take advantage of interactions with a highly collegial group of scientists within NHGRI and on the NIH campus as a whole. In addition, the successful candidate will have access to NHGRI's outstanding core laboratories, as well as the unparalleled resources of the NIH Clinical Center.

Candidates must have a Ph.D., M.D., or equivalent degree, as well as comprehensive, advanced training and a record of accomplishment in one of the targeted areas. This position includes a generous start-up allowance, an ongoing commitment of research space, laboratory resources, and positions for personnel and trainees.

Interested applicants should submit a *curriculum vitae*, a three-page description of proposed research, and three letters of recommendation through our online application system, at http://research.nhgri.nih.gov/apply.

Applications will be reviewed starting Monday, December 15, 2008 and will be accepted until the position is filled.

For more information on GDRB and NHGRI's Intramural Program, please see http://genome.gov/DIR. Specific questions regarding the recruitment may be directed to Dr. William Pavan, the Search Chair, at bpavan@nhgri.nih.gov. Questions may also be directed to Dr. Leslie Biesecker, the GDRB Branch Chief, at leslieb@nhgri.nih.gov.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE   Division of Intramural Research
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

NW172374R

THE NIH IS DEDICATED TO BUILDING A DIVERSE COMMUNITY IN ITS TRAINING AND EMPLOYMENT PROGRAMS

OPPORTUNITIES @ NIH   THE NATIONAL INSTITUTES OF HEALTH

A nice thought

Happy families.

Catherine Mintz

The door was old-fashioned, with a knob you had to turn and push. Harry wondered if he was supposed to be puzzled or show he was knowledgeable by just going in. Even as he pondered, he found he'd opened it. He'd always trusted his intuition.

The office inside was in shades of grey, with three bentwood chairs, an antique plastic wastebasket and nothing else. Not even a place to hang his coat. He took it off, folded it wet side in, and sat down. His shoes were soaked through. He hoped they wouldn't hold it against him that he'd been too excited to check the micro-forecast.

"Mr Lin?" said a voice. "You'll be seen in a moment. If you'd go out the door?"

It was then he realized that there was only one. Harry rose, uncertain. Outside, there was a completely different hall with a red door, invitingly open, opposite. They'd shifted him into the secure section and he'd never felt it.

"Please come in," said the voice. At that, he knew he had succeeded. He would be cloned and his clones optimized for the things he might have done. It was impossible to do as much as he might do in one lifetime.

He would have ten.

A medical droid rolled into the reception area. "Follow me," it said in the voice.

The procedures that had followed were unpleasant but no more so than he had expected.

He had expected more contacts with his clones, which were, after all, his sons, his ten sons, all wonderfully alike and amazingly different, but he accepted that the programme knew what was best for him and for them. He had ten sons, Harry Lins One through Ten. He assigned a finger to each of them and reminded himself of their accomplishments every morning. His contacts with the ten Harrys were brief, but it was important he have their individual identities at his fingertips. He smiled every time he went through his drill: the doctors wouldn't dream he had been so literal.

Harry One was training in biology. It seemed Two would be a physicist. At seven, Harry Three had performed on the violin with the Philharmonic and received a recording contract for his original compositions. "Neither classical nor contemporary, this is music for the ages," read one review of his first album. Harry Four

had at least two possible career paths, poet or actor. Harry Five was a mathematical prodigy. Harry Six …

They were all fine boys and the world would be theirs. He was proud beyond belief. When they attained their majority at 14, he rented a room in the capital's most exclusive restaurant, ordered all their favourite dishes plus a little fine champagne for their first official drink, and memorized a speech about how he loved them all. It was soppy but short. He felt entitled. His investment in them might be emotional, not monetary, but he'd been a good dad.

After dinner, he sat basking in the glow of a life well lived. Harry Ten rose, and said: "Dad, we've been talking things over and I'd like to speak for all of us." Under the table Harry Lin tapped the little finger of his left hand as he remembered that Harry Ten was slow to mature, and had only recently been designated the sociologist. "Experience," the doctors told him, "it's entirely normal for a prodigy in the social sciences to peak later."

"Harry Ten," said Harry Lin and lifted his glass to his son. His sons.

"We've been talking it over, Dad, and we feel that — we mean no offence and hope none will be taken — that you don't add to our image. We're doing well with endorsements and publicity appearances, but we could do better. We will pension you off so that you have a comfortable lifestyle. We thought a simple name change and your signature on this contract not to use ours would be best for all of us. If you'd …"

Harry Lin stared at his children, his sons. Ten faces, each more implacable than the last. The Harry Lin collective could outvote him.

"I'll have my lawyer …"

"He checked it all over and is waiting outside if you'd like to speak with him. Of course, if you prefer another option we can agree to, that would be acceptable. We appreciate all you've done for us." Ten's voiced softened. "Dad, we just want what's best for all of us. That means you, too."

In his mind's eye, Harry, who had all the potential of his ten sons, saw himself as they saw him. Older. Old. Unaccomplished. A blot on the collective's public image. He sat up, acutely aware of an ache in his back that had not bothered him before. "I'll sign," Harry said. "I'll want everything gone over, but I'll sign." He knew he had no choice.

The house he moved to was secluded. However, there was an old-fashioned public cinema nearby where he could watch pre-digital films in their proper context. He started a book on the subject. He'd always wanted to write. Twice a week, a driver ferried him into town for a discreet lunch with one or another of his sons and any necessary business for the collective.

He was coming out of the theatre when a woman, who was of a certain age, stopped him. "You know, you look so much like the Harrys."

"I can't see it myself, but it's a nice thought."

"Would chatting over a drink be a nice thought, too?"

"Yes." He smiled. "Yes, it would. What's your name?"

"Margaret. Margaret Rose."

"Margaret, I'm Harold Lakewood. It's a pleasure to meet you." ∎

Catherine Mintz writes science fiction, fantasy, horror and poetry, both genre and mainstream, as well as non-fiction.

JACEY